Nonprogressive Diffusion on Social Networks: Approximation and Applications

Yunduan Lin
Civil and Environmental Engineering Department, University of California, Berkeley, yunduan_lin@berkeley.edu

Heng Zhang
W. P. Carey School of Business, Arizona State University, hengzhang24@asu.edu

Renyu Zhang
CUHK Business School, The Chinese University of Hong Kong, Hong Kong, China, philipzhang@cuhk.edu.hk

Zuo-Jun Max Shen
Industrial Engineering and Operations Research Department, University of California, Berkeley, maxshen@berkeley.edu

Nonprogressive diffusion describes the dissemination of behavior on a social network, where the agents are allowed to reverse their decisions as time evolves. It has a wide variety of applications in service adoption, opinion formation, epidemiology, etc. Building upon the rich studies in network diffusion analysis and operations research, we propose a general model to characterize nonprogressive diffusion and develop a fixed-point approximation (FPA) scheme to characterize the limiting adoption on a social network. This approximation scheme admits both a theoretical guarantee and computational efficiency. We show that the maximal deviation of the FPA scheme diminishes as the network size and density increase at a rate of $\mathcal{O}(1/\sqrt{N_{\min}})$, where $N_{\min}$ is the minimum indegree of the agents on a social network. Thus, the FPA scheme is most powerful for dense and large networks that are generally prohibitive by simulation. Taking the widely studied influence maximization and pricing problems on a social network as examples, we further illustrate the broad applications of our FPA scheme. Finally, we conduct comprehensive numerical studies with synthetic and real-world networks. The FPA scheme shows 1,000 times speed up in computation time than simulation. It achieves small approximation error, and outperforms conventional algorithms even when the social network is small and/or sparse.

Key words: Nonprogressive network diffusion, Large-scale network approximation, Influence maximization, Pricing

1. Introduction

Social networks fundamentally shape our lives. People are more receptive to information shared by their friends and relatives (Acemoğlu et al. 2013, Lu et al. 2013) and more inclined to make a purchase when informed by their acquaintances (Ma et al. 2015, Bapna and Umyarov 2015). It is even more so in the digital era—globally, 4.62 billion people, approximately 58.4% of the population worldwide, used online social network platforms, such as Facebook, YouTube and Tiktok, by January 2022 (Datareportal 2022). Such online platforms expand the reach of our social networks. Friends and strangers online also influence our opinions and decisions. As a result, the prevalence of such online platforms provides unprecedented opportunities for firms to exploit the social network of their customers to improve their marketing or operations strategies.

In general, each agent on a social network is influenced by, and influences others, simultaneously. As a consequence, technology innovations, political campaigns, and breaking news all diffuse widely on social networks. In practice, platforms that leverage such diffusions on social networks have the chance to enhance their impact and gain more profit (Shriver et al. 2013, Dewan et al. 2017). On the other hand, the diffusion on social networks is very complicated because it is a combined effect of agents' personal characteristics and social relationships. To understand the mechanisms of diffusion in networks, network diffusion analysis has been a long-lasting topic that has appealed to generations of researchers (see books Jackson 2010, Shakarian et al. 2015).

Network diffusion analysis spans multiple fields of study, such as computer science (Kempe et al. 2003, Acemoğlu et al. 2013), economics (Sadler 2020), operations management (Song and Zipkin 2009, Candogan et al. 2012, Shen et al. 2017, Wang and Wang 2017), epidemiology (Kermack and McKendrick 1927, Drakopoulos and Zheng 2017) and sociology (Wasserman 1994, Lewis and Kaufman 2018). Due to the diverse application domains, there is not a one-size-fits-all approach to model the diffusion process. As categorized by the seminal paper Kempe et al. (2003), there are two types of diffusion processes in general, progressive diffusion and nonprogressive diffusion. For progressive diffusion, an agent can *only* switch from the unadopted status to the adopted status but not vice versa. The progressive diffusion model applies to settings with a single-directional status change, such as adopting new technology, consuming online content, purchasing a product, etc. Our study, instead, focuses on the nonprogressive diffusion case where a state transition can occur in both directions. Nonprogressive diffusion processes are prevalent in practice. Wide applications can be found in the settings where the adoption decision can be withdrawn or reset, such as choosing a cell phone service provider, voting for a political figure, signing up for a membership program, subscribing a YouTube channel, being infected in a pandemic, etc.

Broadly, there are two approaches to the study of diffusion in social networks. The first approach is more microfounded by capturing the concrete network topology and the dynamic stochastic evolution of agent states. This stream of study includes, the independent cascade model (Goldenberg et al. 2001) and the linear threshold model (Granovetter 1978, Schelling 1978). These models provide accurate descriptions of the diffusion patterns at the individual agent level over time. However, such a refined model is at the cost of technical tractability—simulation happens to be the only technique for its analysis. In particular, one needs to repeat tens of thousands of simulation runs to measure the influence of diffusion, which is a computationally inefficient exercise. As a consequence, the follow-up optimization problem, even for a sparse and moderate-sized network, is time-consuming, if not intractable (Chen et al. 2009). Notwithstanding the plentiful literature on these models, we still lack good tools to analyze them efficiently. The other approach simplifies the diffusion process by either assuming away the network topology and focusing on the state of the entire population (e.g., SIR/SIS and Bass models; Kermack and McKendrick 1927, Bass 1969, Lin et al. 2021), or ignoring the state evolution and focusing on the equilibrium outcome (Candogan et al. 2012, Jackson et al. 2020). These models focus on the equilibrium behavior characterization of network diffusion without specifying how the equilibrium is reached, which greatly facilitates the subsequent analysis to provide sharp insights.

Our work bridges these two approaches by providing a simple, efficient and accurate approximation scheme for the nonprogressive diffusion processes in a general social network. The proposed approach not only provides a simple and efficient characterization of the diffusion outcome but also leads to tractable solution techniques for optimizing diffusion. We build a general nonprogressive dynamic diffusion model encompassing the following properties. (1) *Microfoundedness.* We characterize the microscopic behavior of each agent in a discrete-time fashion, which allows us to characterize the heterogeneous effect of diffusion on every single agent in the social network. (2) *Local network effect.* An agent's adoption decision is locally influenced by his or her neighbors. In other words, the network effects influencing each agent are also heterogeneous. (3) *General network topology.* We do not impose any restrictions on the network structure. These three properties ensure the practicality of our model and the wide applicability of our analysis technique and results.

Although characterizing the long-run adoption rate for each agent in a model with such granularity does not seem technically tractable, we develop a fixed-point approximation (FPA) scheme that estimates the adoption rates through a set of easily solvable fixed-point equations. We demonstrate that the proposed FPA scheme is not only computationally efficient but also admits a provable guarantee that ensures that its maximal deviation from the limiting adoption is small for sufficiently large and dense networks. Our approximation scheme further paves the way for optimizing operational decisions, such as the influence maximization and pricing problems in a social network. One can easily formulate such problems and design computationally efficient algorithms that provide near-optimal solutions. In summary, through the FPA scheme, we show that the diffusion outcome characterized by our "micromodel" can be approximated by an easy-to-analyze "macromodel." Therefore, our method enjoys the advantage of both modeling approaches.

1.1. Contributions and Organization

Our research contributions are summarized as follows:

- **General modeling framework and approximation scheme.** We analyze nonprogressive diffusion with a dynamic, stochastic process, a general modeling framework that embraces many of the diffusion models in the literature. Our model is microfounded, capturing the local network effect and individual heterogeneity with no requirements on the network structure. To the greatest extent possible, our model is flexible and general enough to incorporate both the detailed social network information and the features of each agent. Under this framework, we propose the FPA scheme for the limiting adoption probability of each agent on the social network. We develop a nontrivial "fixed-point sandwich" proof technique to show that the maximal deviation between the FPA scheme and the exact adoption probability of any agent is upper bounded by $\mathcal{O}(1/\sqrt{N_{\min}})$, where $N_{\min}$ is the minimum indegree of all the agents on the network (Theorem 1). Therefore, as $N_{\min}$ increases (i.e., When the social network becomes larger and denser), the FPA scheme is asymptotically equivalent to the limiting adoption probability. In other words, our approximation is the most powerful for dense and large networks that are generally prohibitive to handle by simulation.

Through comprehensive numerical studies on both synthetic and real-world networks, we find that the FPA scheme is much more effective than the theoretical bound for most of the problem instances we examine. In particular, it provides an accurate approximation for the limiting adoption rates for small and sparse networks where the indegree of any agent is small. For example, as highlighted in the numerical experiments, when agents only have an average degree of 10 in the social network, the FPA scheme achieves a mean absolute percentage error (MAPE) of less than 10%. Furthermore, the numerical results also highlight the superior efficiency of our approach over simulation, with the former being at least 1,000 times faster.

- **A wide applicability in optimizing operational decisions.** The FPA scheme is powerful enough to address applications in multiple fields. In particular, we can reformulate operational decision-making problems using the FPA scheme, which, by virtue of our approximation error bound (Theorem 1), will lead to high-quality decisions in optimizing the diffusion outcome. Such reformulation results in solution techniques for diffusion problems were previously too difficult to solve. We take two well-known operational problems, influence maximization (IM) and pricing on a social network, as examples to demonstrate this wide range of applications. For the IM problem, we show that under mild technical conditions, the influence function is submodular with regard to the seed user set. This extends the classical simulation-based greedy algorithm to more general diffusion settings. More importantly, it performs magnitudes more efficiently than such simulation-based methods with virtually no performance loss. Large-scale numerical experiments also demonstrate that the greedy algorithm works very well on nonsubmodular problems.

For the pricing problem on a social network, the FPA scheme provides an easy-to-optimize static reformulation. Indeed, it seems that the only alternative method would be executed through a cumbersome simulation-based numerical gradient descent. Through FPA, we provide near-optimal algorithms to determine the prices for profit maximization. Specifically, we show that, under technical conditions, the price optimization problem is a convex program in the adoption probability space, as long as the platform can perfectly price discriminate each agent. In more general settings where such technical assumptions are lifted, we derive approximate gradient expressions for the direct optimization in the price space. We show that this achieves near-optimal solutions efficiently for settings such as pricing with uniform prices.

The remainder of the paper is structured as follows: In Section 2, we review the related literature. Section 3 introduces the nonprogressive diffusion model and characterizes the limiting adoption rate of each agent. In Section 4, we propose the FPA scheme for limiting adoption probability and demonstrate its theoretical and numerical performances. We study the IM problem and pricing problem on a social network using our FPA scheme in Section 5. Section 6 concludes this paper.

2. Literature Review

Our paper is broadly related to the literature on network diffusion. We first briefly review network diffusion models in different settings. Then, we discuss some operational optimization problems that involve network diffusion.

2.1. Network Diffusion Model

Diffusion in the social network is an attractive topic in many domains. Researchers have proposed different models for specific applications: Epidemiologist use the SIR model (Kermack and McKendrick 1927) to study the propagation of epidemics; Marketing researchers introduce the Bass model (Bass 1969) for the diffusion of new innovations; Computer scientists propose the linear threshold model (Granovetter 1978, Schelling 1978) and independent cascade model (Goldenberg et al. 2001) for data mining. Although these models are disparate, the one thing they have in common is the more there is knowledge known about the network and agents, the more complicated the model is. Subsequently, much more demanding operational problems have arisen. As an example, the Bass model ignores most information on network structure and agents but enjoys the advantages of the analytical expressions on some critical values, including the number of cumulative adopters and the time at which the adoption rate reaches the peak. Consequently, it is easy to derive an optimal pricing policy under a Bass-type model (Robinson and Lakhani 1975, Agrawal et al. 2021). In contrast, the linear threshold model incorporates the entire network structure, and Chen et al. (2010) proves that it is #P-hard in general to compute its diffusion influence. Hence, the follow-up works on the linear threshold model (Goyal et al. 2011, Tang et al. 2014) rely heavily on simulation techniques. In conclusion, a natural trade-off arises when people want to characterize diffusion, that is, they either trade the concreteness of the diffusion for practical efficiency or vice versa.

In addition to the classical diffusion models, there are some parallel streams of work involving network diffusion and relating to our work. First, a variety of engineering and economics applications describe the interactions across the network using network games (Ballester et al. 2006, Jackson and Yariv 2007, Candogan et al. 2012, Afeche et al. 2018, Baron et al. 2022, Feng et al. 2022). A central goal of the literature on network games is to analyze various types of equilibria. Although our fixed-point approximation is reminiscent of the equilibrium in the network games, the motivation and analysis are entirely different. Network games focus more on strategic behavior, where agents adopt the best strategy in response to others' strategies. Our model does not require the agents to be strategic. Instead, we capture the dynamics and stochasticity of tactical agents, where agents only need to make decisions according to the observation of real-time actions from others. Second, a number of operations management studies incorporate network externality into consumer choice models. This type of work, serving for the subsequent assortment or pricing problem, generally simplifies the knowledge of the network structure. Du et al. (2016), Wang and Wang (2017) consider the global effect where the agent is affected by the average adoption of the entire market. Nosrat et al. (2021) segments the network into multiple segments and consider the effects between these segments. Gopalakrishnan et al. (2022) studies the local network effect by taking the degree of each agent as a proxy. Xie and Wang (2020) considers the local network effect but only restrict the analysis to the star network and directed acyclic graph. Our work, instead, fully considers the network information and embraces a more general setting. Finally, our work also closely relates to an extensive array of works studying the mean-field approximation for stochastic processes (Benaïm and Weibull 2003, Van Mieghem et al. 2008, Sridhar and Kar 2021). While mean-field approximation is essentially descriptive, generally providing a deterministic approximation at the population level, our work focuses more on the operational side of each agent individually.

2.2. Optimization Problems with Network Diffusion

The FPA scheme is applicable in a wide variety of network diffusion applications. In our work, we demonstrate two examples, influence maximization and pricing problems on a social network. Therefore, we mainly review the literature on these two applications.

Kempe et al. (2003) first consider the issue of choosing an influential set of agents to maximize the spread of diffusion influence as a problem in discrete optimization. They show that for a large class of diffusion models, the IM problem is NP-hard, and the objective is a submodular function with regard to the influential set. Therefore, the most widely used approximation algorithm for the IM problem is the greedy algorithm of submodular maximization. However, to find K seed users in an N-node network, the greedy algorithm requires evaluating the diffusion influence for

NK different influential sets. As the evaluation of each influential set is #P-hard, it is time-consuming to achieve the approximated solution. Other existing approaches can be classified into three categories, namely, simulation-based (Leskovec et al. 2007, Goyal et al. 2011, Zhou et al. 2015), sketch-based (Cheng et al. 2013, Tang et al. 2014), and proxy-based (Chen et al. 2010, Wang et al. 2012). We refer readers to the survey (Li et al. 2018) for a more comprehensive review. However, these approaches sacrifice either accuracy or efficiency and are not ideal for practical use. With the FP approximation we proposed in this work, we can take both into consideration.

For the pricing problem, there is a growing literature in the economics and operations management communities that considers the presence of network effects (Xie and Sirbu 1995, Anari et al. 2010, Hu et al. 2020, Li 2020, Yang and Zhang 2022). Recent studies on the single-item pricing problem with the network effect can be found in Candogan et al. (2012), Du et al. (2018) and Nosrat et al. (2021). In the following, we go through the main ideas of these three papers and explain the differences with our paper. Candogan et al. (2012) use a network game to characterize the agent behavior where the utility function follows a quadratic form. Our work, instead, considers a more general setting where common utility models can all be applied. Du et al. (2018) considers a scenario where there are multiple equilibria, and the price cannot be uniquely determined. Nosrat et al. (2021) study a pricing problem where there are multiple segments of customers. These two papers do not consider different pricing schemes, and restrict the platform to offer a uniform price. Our formulation, as well as the proposed algorithms, can be applied in a more flexible setting.

3. Nonprogressive Network Diffusion Model

In this section, we first introduce our network diffusion model and then characterize the limiting behavior of each agent in the network. Our model can be applied to a wide range of nonprogressive diffusion settings, among which we use service adoption on an online social network platform as the context for illustration.

3.1. Preliminaries and Formulation

We model the social network platform (e.g., TikTok) as a graph $G=(V,E)$, where $V:=\{1,2...,|V|\}$ is the set of all agents and $E:=\{1,2...,|E|\}$ is the set of all directed edges. A directed edge $(i,j)\in E$, where $i,j\in V$, implies that agent i is influenced by agent j, and we call j an *in-neighbor* of i. We interpret $(i,j)\in E$ as i following j on the social network. We use $\mathcal{N}_i$ to denote the set of all in-neighbors for agent i (i.e., $\mathcal{N}_i:=\{j\in V:(i,j)\in E\}$) and $n_i:=|\mathcal{N}_i|$ to denote her indegree (i.e., the number of in-neighbors).

We use t to denote the discrete time period, and $t=0$ refers to the service launching time. Define $Y_i(t)\in\{0,1\}$ as the state of agent i at period t, where $Y_i(t)=1$ (resp. $Y_i(t)=0$) means the adoption (resp. nonadoption) of the service in this period. Without loss of generality, we normalize the utility of nonadoption to 0. The initial state of the agents $\boldsymbol{Y}(0)$ follows an arbitrary distribution on $\{0,1\}^{|V|}$. For all $t \geq 1$, each agent $i \in V$ will decide whether to adopt the service based on her realized utility in period t: $Y_i(t) := \mathbb{1}\{u_i(t) \geq 0\}$, where $u_i(t)$ is the utility of agent i to adopt the service in period t, and is defined as follows:

$$u_i(t) := v_i + \beta \cdot \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{n_i} + \epsilon_i(t) \tag{1}$$

As shown by Eqn. (1), the utility of agent i in period t to adopt the service consists of three parts: (a) the idiosyncratic intrinsic value v_i, (b) the local network effect $\beta \cdot \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{n_i}$ and (c) the random noise $\epsilon_i(t)$. The idiosyncratic value v_i reveals the personalized preference of the agent and remains unchanged throughout the horizon. From an analytical point of view, with the support of big data, the decision-maker can fully "featurize" it. In other words, v_i can be estimated with agent features such as her demographic information and behavioral data, and is sometimes affected by the platform's operational strategies. For example, the price of a paid service (e.g., YouTube Premium) will definitely affect whether and how, the agent likes the service. The local network effect (i.e., the second term of Eqn. (1)) captures the peer influence on the agent from her in-neighbors through their average adoption rate in the previous period, with parameter β to quantify the network effect intensity. If an agent i does not have in-neighbors (i.e., $\mathcal{N}_i = \emptyset$), we set this term to 0. Finally, we assume the random noise $\epsilon_i(t)$ is independent and identically distributed (i.i.d.) across agents and time periods. We assume, without loss of generality, that $\mathbb{E}[\epsilon_i(t)] = 0$. We do not put any restrictions on the distribution of $\epsilon_i(t)$ except for the following mild condition in Assumption 1.

Assumption 1 **(Lipschitz Continuity)**. *The noise $\epsilon_i(t)$ has an L-Lipschitz continuous CDF: $|F_\epsilon(x) - F_\epsilon(y)| \leq L|x-y|$ for any $x, y \in \mathbb{R}$.*

Therefore, we require that the noise distribution is sufficiently smooth. Assumption 1 is satisfied by any continuous distribution with a bounded probability density function. Common distributions, such as logistic distribution and normal distribution, satisfy these conditions and thus can be handled by our model. We also impose an upper bound for the network effect parameter β that facilitates characterizing the limit of diffusion.

Assumption 2 **(Bounded Network Effects)**. *The network effect satisfies $|\beta L| < 1$.*

Parameter β quantifies the magnitude of network externality. Similar assumptions are commonly made in the network economics literature (e.g., see Horst and Scheinkman 2006, Wang and Wang 2017, Xu 2018, Jackson et al. 2020, Gopalakrishnan et al. 2022). In such settings, they are often introduced to ensure that the network effect is not too substantial, and as a result, the equilibrium in a game-theoretical sense is well defined. However, as we will discuss later, this assumption not only excludes divergent or periodic behavior in the long run in our diffusion model (Theorem 2) but also guarantees a simple fixed-point characterization of the limiting adoption probabilities (Theorem 1). When $\beta > 0$, the network exhibits a positive externality; users have a higher adoption utility when observing more neighbors adopted. When $\beta < 0$, the network, in contrast, exhibits a negative externality. Our subsequent analysis carries through if $-\frac{1}{L} < \beta < 0$. For simplicity, we assume, without loss of generality, that the network effects are positive and $0 < \beta < \frac{1}{L}$ throughout.

A natural objective of network diffusion analysis is to quantify the total diffusion. For example, some previous studies (Kempe et al. 2003, Chan et al. 2020) used the adoption frequency throughout a fixed time horizon to denote the influence. In this work, following the literature, we focus on a natural metric, the limiting adoption probability in the long run. Provided it converges, it also represents accumulated reward (number of adoptions) in the long run.

We remark on the notations. Hereafter, we use a bold math notation to denote the collection of a particular variable over all the agents in vector form. We note that the network structure, as well as the intrinsic values, determine how diffusions evolve. Thus, when we refer to a specific diffusion instance, we are given a fixed network G and intrinsic value $\boldsymbol{v}$; when we refer to a sequence of diffusion instances, we are given a sequence of pairs $(G, \boldsymbol{v})$.

3.2. A Markov Chain Perspective

Notably, each diffusion instance can be characterized by a Markov chain (MC), of which the state space is the set of indicator vectors denoting all the possible combinations of adoption decisions $\{0,1\}^{|V|}$. The one-step transition probability can be calculated based on the conditional probability shown in Eqn. (2).

$$\begin{aligned}\mathbb{P}(Y_i(t)=1|\boldsymbol{Y}(t-1)=\boldsymbol{y}) &= \mathbb{P}\left(v_i + \beta\frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{n_i} + \epsilon_i(t) \geq 0 \middle| \boldsymbol{Y}(t-1)=\boldsymbol{y}\right)\\ &= 1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} y_j}{n_i}\right)\end{aligned} \tag{2}$$

Consequently, the transition probability from state $\boldsymbol{y}$ to $\boldsymbol{y}'$ can be represented by

$$P(\boldsymbol{y},\boldsymbol{y}') = \prod_{i\in V} F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} y_j}{n_i}\right)^{1-y_i'} \cdot \left[1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} y_j}{n_i}\right)\right]^{y_i'}$$

As we are concerned with the overall adoption probability of an agent rather than the probability of a single MC state, we define the adoption probability of agent i at time t as Eqn. (3).

$$q_i(t) := \mathbb{P}(Y_i(t)=1) \equiv \sum_{\boldsymbol{y}\in\{0,1\}^{|V|}} \mathbb{1}\{y_i=1\}\cdot\mathbb{P}(\boldsymbol{Y}(t)=\boldsymbol{y}) \tag{3}$$

We have the following proposition on the limiting behavior of $\boldsymbol{q}(t)$ when t tends to infinity.

Proposition 1 **(Limiting Adoption Probability)**. *Under Assumption 1 and Assumption 2, the adoption probability of each agent i converges to*

$$\lim_{t\to\infty} q_i(t) = q_i^* := \sum_{\boldsymbol{y}\in\{0,1\}^{|V|}} \mathbb{1}\{y_i = 1\}\cdot \pi(\boldsymbol{y})$$

when t increases, where $\boldsymbol{\pi}$ is the stationary distribution of the MC that satisfies $\boldsymbol{\pi} = \boldsymbol{\pi} P$.

As made clear in the proof of Proposition 1, with our assumptions on random noise and network coefficient, the MC only has a *single aperiodic recurrent class.* Thus, a limiting distribution $\boldsymbol{\pi}$ leads to the limiting adoption probabilities $\boldsymbol{q}^*$ for all agents. Furthermore, by the standard MC theory, one can easily verify that

$$\lim_{t\to\infty} \frac{1}{t}\cdot\sum_{s=1}^{t} Y_i(s) = q_i^* \text{ a.s.} \quad \text{and} \quad \lim_{t\to\infty} \frac{1}{t}\cdot\sum_{s=1}^{t} q_i(s) = q_i^* \;\forall i\in V. \tag{4}$$

As a result, on the operational side, Proposition 1 allows us to utilize the limiting adoption probability $\boldsymbol{q}^*$ to formulate a network diffusion related optimization problem. Formally, a wide range of operational problems involving network diffusion, such as the influence maximization problem (Section 5.1) and the pricing problem on a social network (Section 5.2), can be formulated as the following optimization problem:

$$\max_{(G,\boldsymbol{v})\in\mathcal{I}} f(\boldsymbol{q}^*(G,\boldsymbol{v}),\boldsymbol{v}), \tag{5}$$

where $\mathcal{I}$ denotes a set of feasible diffusion instances incorporating operational decisions (e.g., targeted seed users and prices), and $f(\cdot)$ is the objective function measured on the limiting adoption probability. With a slight abuse of notation, we use $\boldsymbol{q}^*(\cdot)$ to denote the mapping from diffusion instance $(G,\boldsymbol{v})$ to the limiting adoption probability vector characterized by Proposition 1. For the influence maximization problem, $\mathcal{I}$ refers to the set of diffusion instances where the intrinsic values of a set of agents (i.e., seed users) are set to be arbitrarily large values, and $f(\cdot)$ represents the total expected limiting adoptions. For the pricing problem on a social network, $\mathcal{I}$ includes diffusion instances with different intrinsic values affected by a range of pricing schemes, and $f(\cdot)$ is the profit function given the price and limiting adoption probabilities.

However, the optimization (5) is usually difficult to solve due to the absence of closed-form expressions for $\boldsymbol{q}^*$. Indeed, the state number of MC is $2^{|V|}$, which is exponential in the network size. It is intractable to construct the transition matrix even for a moderate-sized network, let alone to calculate the limiting adoption probabilities $\boldsymbol{q}^*$. Therefore, the optimization problem (5) is generally intractable both analytically and computationally, which motivates us to develop our approximation scheme for $\boldsymbol{q}^*$ presented in Section 4.

3.3. Connections to Other Diffusion Models

Before presenting our approximation scheme, we comment on how our diffusion model is related to a variety of models in the network diffusion literature.

- *Linear threshold (LT) model.* Our model is closely related to the LT diffusion model introduced in (Granovetter 1978). It is one of the most basic and widely studied diffusion models, and is commonly used to describe the spread of influence, and the cascading behavior in network diffusions (Kempe et al. 2003, Chen et al. 2010). In the LT model, each agent is associated with a random threshold i.i.d. sampled from the uniform distribution $\mathcal{U}(0,1)$ and will become adopted when the network effect exceeds this threshold. We extend the classical LT model by introducing the deterministic intrinsic values of the agents. To be specific, one can view the classical LT model as a special case of our model where all agents have the same intrinsic value $v_i = -0.5$, and the random noise $\epsilon_i(t)$ follows the uniform distribution $\mathcal{U}(-0.5, 0.5)$. In fact, our model decomposes the random threshold in the LT model into a known deterministic intrinsic value and random noise. Thus, we allow the heterogeneity of users' organic preferences and different types of random noise. One advantage of LT models is that they accurately describe rational agents' decision processes and hence the evolution of the diffusion processes. They provide a solid microfoundation to model diffusion, however, except for relying on simulation, it seems next to impossible to characterize the diffusion processes in such models. Due to the sheer scale of the networks in many applications, simulation often becomes cumbersome. Our approach provides a remedy for this issue.
- *Discrete choice model with network effects.* Discrete choice models describe choices between several discrete alternatives. Previous literature (Du et al. 2016, Wang and Wang 2017, Gopalakrishnan et al. 2022) usually models the user behavior in a network at each time step as a binary or multinomial logit model, a type of discrete choice model. One may consider our work utilizing a choice model as restricted to two or binary choices. We remark on the two aspects. First, our model covers a larger collection of discrete choice models than the standard logistic distributional assumption on random noise. For example, when the random noise follows a uniform and normal distribution, our model aligns with the linear probability model and probit model accordingly. Second, we include the stochasticity of the network effect at each time step into the model by assuming that the network effect comes from realized average adoptions, as opposed to the expected adoption rate commonly assumed in the literature. However, interestingly, our main result (Theorem 1) shows that in the long term, an expected adoption rate treatment rising naturally from a game-theoretic perspective (Du et al. 2016, Wang and Wang 2017, Gopalakrishnan et al. 2022) is in fact well-justified in a dynamic and microfounded setting. In other words, our work bridges the LT models and these game-theoretical choice models. We also mention that we present a binary choice model for ease of analysis but our approach can be extended to settings with more choices.

- *Model with global network effect.* The global network effect is generally used when the network structure cannot be observed or when agents are assumed to have access to the actions of all others (Nosrat et al. 2021, Chen and Chen 2021). The global network effect assumes that adoption is affected only by the fraction of adopters in the entire population, which can be captured by our model with a complete underlying network.

4. Fixed-Point Approximation

In this section, we provide a fixed-point approximation (FPA) scheme for the limiting adoption probability of all the agents. We analyze the quality of the FPA scheme on different diffusion instances, and numerically show the performance on extensive synthetic and real-world networks. In this section, we assume that Assumptions 1 and 2 hold.

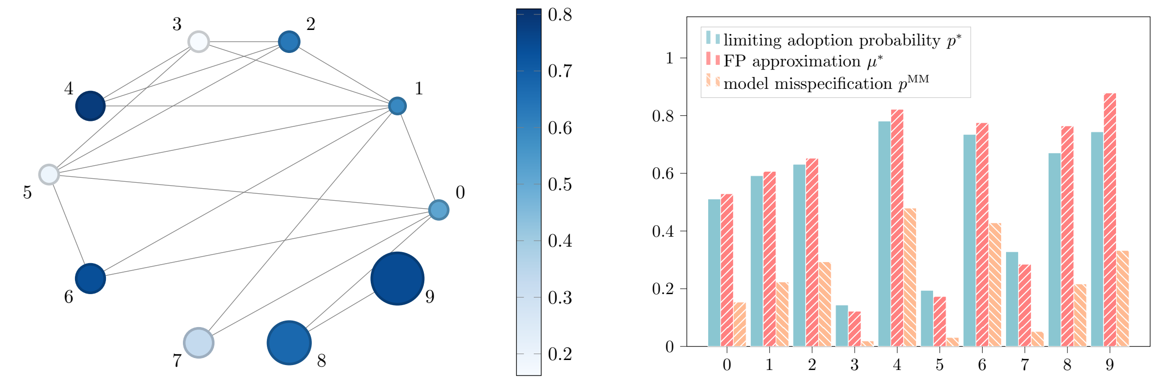
4.1. Overview and Motivating Example

For a given diffusion instance with network G and intrinsic value $\boldsymbol{v}$, we will show that the limiting adoption probability $\boldsymbol{q}^*$ can be reasonably approximated by the solution $\boldsymbol{\mu}^*$ of the following simple system of equations.

$$\mu_i = 1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i}\mu_j}{n_i}\right) \text{ for all } i \in V \tag{6}$$

We first use an example of a 10-node network to demonstrate the results of the limiting adoption probability $\boldsymbol{q}^*$ under the diffusion model and its FPA solution $\boldsymbol{\mu}^*$. In this diffusion instance, the characteristics of heterogeneity exist in both the network degrees and the intrinsic values among the agents. We obtain $\boldsymbol{q}^*$ by simulation and the FPA solution $\boldsymbol{\mu}^*$ by solving (6), respectively. Moreover, for a better intuition of the overall diffusion influence, we also calculate the adoption probability under a model misspecification where the network effect is not considered, that is, $q_i^{\mathrm{MM}} = \mathbb{E}[\mathbb{1}\{v_i + \epsilon_i \geq 0\}]$. In Appendix B.1, we include detailed information about this instance as well as the numerical results.

Figure 1a illustrates the network structure and exhibits the approximation results. Specifically, for node i, the color denotes the true value of q_i^*, and the size denotes the absolute error $|q_i^* - \mu_i^*|$. Comparing different nodes, we notice that the absolute error is relatively large when the node has fewer neighbors (i.e., nodes 8 and 9). In contrast, nodes that connect to many other nodes (i.e., nodes 0, 1 and 5) have smaller errors. Figure 1b further shows the values of $\boldsymbol{q}^*$, $\boldsymbol{\mu}^*$ and $\boldsymbol{q}^{\mathrm{MM}}$. The network effect evidently affects the adoption probabilities but the FPA scheme captures it accurately. Taking $\boldsymbol{q}^*$ as a baseline, the mean absolute error values for $\boldsymbol{\mu}^*$ and $\boldsymbol{q}^{\mathrm{MM}}$ are 0.045 and 0.306, respectively. This example provides suggestive evidence of the high quality of the FPA scheme and hints that the approximation scheme seems to favor agents with more neighbors.



(a) Approximation results on the network structure

(b) Value of adoption probability

Figure 1 An example to illustrate the fixed-point approximation. Left: Approximation results on the network structure (The color denotes the true value of limiting adoption probability. The size denotes the absolute error between the FPA solution and the limiting adoption probability); Right: Value of adoption probability.

In the following, we show that under Assumptions 1 and 2, for any diffusion instance, there always exists a unique solution $\boldsymbol{\mu}^*$ to (6). More importantly, we show that the maximal deviation between $\boldsymbol{q}^*$ and $\boldsymbol{\mu}^*$ shrinks when the diffusion instance is associated with a denser and/or larger network. The minimum indegree of network $N_{\min} = \min_{i \in V} n_i$ is a critical indicator in this result. See Theorem 1 below for the formal statement of this result. The key technical challenge here is that the adoption decisions are highly correlated across the network, and as time evolves. We introduce a "fixed-point sandwich" proof technique to show that the approximation error diminishes when the network becomes larger and denser. Since the FPA solution $\boldsymbol{\mu}^*$ can be directly derived by performing a fixed-point iteration scheme (Rheinboldt 1998), which converges in linear time, we can efficiently compute the FPA solution $\boldsymbol{\mu}^*$.

4.2. The Approximation Error

We now formally present a bound on the error of the FPA scheme. We use the maximal deviation $\|\boldsymbol{q}^* - \boldsymbol{\mu}^*\|_\infty$ as a metric, which measures the L^∞-norm of the difference between $\boldsymbol{q}$ and $\boldsymbol{\mu}^*$ and represents the largest deviation of the FPA solution from the true limiting adoption probability among all the agents. Our key result is the following theorem.

Theorem 1 **(The Error of the FPA Scheme)**. *For any diffusion instance with minimum indegree $N_{\min}$, the maximal deviation of the fixed-point proxy can be upper bounded by*

$$\|\boldsymbol{q}^* - \boldsymbol{\mu}^*\|_\infty \le C\sqrt{\frac{1}{N_{\min}}},$$

where $C = \frac{2}{1-L\beta}\sqrt{\frac{(L\beta)^2}{1-(L\beta)^2}}$.

Theorem 1 shows that for a sequence of diffusion instances with an increasing minimum indegree $N_{\min}$, the maximal deviation shrinks at the rate of $\mathcal{O}\left(\sqrt{1/N_{\min}}\right)$. When $N_{\min}$ increases to infinity, the FPA solution $\boldsymbol{\mu}^*$ is asymptotically equal to the adoption probability $\boldsymbol{q}^*$. This result justifies the evidence from the example in Section 4.1. The approximation performance is strongly related to the number of neighbors while insensitive to the intrinsic values. To upper bound the maximal deviation among all the agents, $N_{\min}$ plays an important role. This is because, although there might be only a few agents having a small number of neighbors, the error will accumulate and spread across the entire network as time evolves so that all the agents in the network may be affected.

The significance of the FPA solution $\boldsymbol{\mu}^*$ lies in the fact that it allows us to reformulate and simplify the optimization (5) for a variety of operational contexts. Instead of solving (5) directly, we replace the exact limiting adopting probabilities $\boldsymbol{q}^*$ with the approximated values μ^* and approximate (5) as follows.

$$\max_{(G,\boldsymbol{v})\in\mathcal{I}} f(\boldsymbol{\mu}) \quad \text{s.t. } \boldsymbol{\mu} = \boldsymbol{h}(\boldsymbol{\mu}\,|\,G,\boldsymbol{v}), \tag{7}$$

where $\boldsymbol{h}(\cdot\,|\,G,\boldsymbol{v})$ is the adoption evolution operator induced by the diffusion instance $(G,\boldsymbol{v})$, which we define formally by Eqn. (8) in Section 4.3.

We advocate the approximate optimization problem (7) in that the FPA scheme brings great benefits to the network diffusion analysis. The advantages are threefold: (i) *Theoretical guarantee.* The asymptotic equivalence makes the FPA solution $\boldsymbol{\mu}^*$ appealing, especially for large, dense networks. From a mathematical perspective, simple theoretical results can be obtained for typical network structures, such as random graphs, regular networks and complete networks, when the number of nodes or edges increases (see the experiments on Erdős–Rényi random graphs in Section 4.4.1). From an application perspective, real-world networks are usually large scale and will continue to grow with more connections. The approximation quality of the FPA solution improves as the network evolves. In addition, although sometimes the real-world network might be sparse, the active users that attract the major interest in the platform usually have many connections (Ghosh and Lerman 2012). Therefore, it seems promising to use the FPA scheme in real-world analyses (see the experiments on two real-world networks in Section 4.4.2). (ii) *Computational efficiency.* The computation of the FPA solution is much faster than the traditional Markov chain Monte-Carlo (MCMC) simulation techniques. According to Rheinboldt (1998), a one-time fixed-point iteration that converges in linear time can directly lead to the FP solution. In contrast, MCMC requires conducting a long-length simulation run, which is a realization of the adoption procedure under stochasticity. The limiting adoption probability is then estimated by the average adoption times over a long period. Furthermore, a large network needs an even longer simulation run to achieve a comparable maximal deviation with the FPA scheme. (iii) *Closed-form expression.* As shown in Eqn. (6), the FPA solution $\boldsymbol{\mu}^*$ can be captured by a closed-form formulation. On the one hand, the FPA scheme is a simple expression for the complex diffusion model. On the other hand, for the subsequent operational decision problems, it is much easier to deal with Eqn. (7) compared with Eqn. (5). Thus, it has the potential to develop more efficient algorithms for these problems. We illustrate the potentially wide applicability of our FPA scheme via its applications to the influence maximization problem and the pricing problem on a social network in Sections 5.1 and 5.2, respectively.

4.3. Details of Error Analysis

In this part, we elaborate the analysis that leads to Theorem 1. As our main methodological contribution, the key idea of our analysis is to construct a fixed-point approximation version of the diffusion process $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ for each instance (G, v), to which the adoption probability process $\{\boldsymbol{q}(t)\}_{t=0}^{\infty}$ is sufficiently close. As a consequence, $\boldsymbol{q}^*$, as the limit of $\boldsymbol{q}(t)$, should be close to $\boldsymbol{\mu}^*$, which is the limit of the approximated diffusion process $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$. To bound the distance between $\{\boldsymbol{q}(t)\}_{t=0}^{\infty}$ and $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$, we establish that the intratemporal and intertemporal correlations of the agent adoption states $\{\boldsymbol{Y}(t)\}_{t=0}^{\infty}$ are both within an order of $\mathcal{O}(1/\sqrt{N_{\min}})$, which guarantees the same order of $||\boldsymbol{q}^* - \boldsymbol{\mu}^*||_{\infty}$.

Specifically, we define the approximated diffusion process $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ as a deterministic dynamical system by Eqn. (8) throughout the entire time horizon:

$$\mu_i(t) = \begin{cases} q_i(0) & t = 0 \\ 1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} \mu_j(t-1)}{n_i}\right) & t > 0 \end{cases}, \text{ for all } i \in V \tag{8}$$

We let $\boldsymbol{h} : \mathbb{R}^{|V|} \to \mathbb{R}^{|V|}$ be the mapping such that the dynamical system $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ can be written as $\boldsymbol{\mu}(t) = \boldsymbol{h}(\boldsymbol{\mu}(t-1))$ for $t \geq 1$. We call $\boldsymbol{h}(\cdot)$ the adoption evolution operator (AEO) hereafter. To assist the analysis, we also define a family of auxiliary AEOs

$$\mathcal{H} := \left\{\boldsymbol{h}_\zeta(\cdot) = \boldsymbol{h}(\cdot) + \zeta \boldsymbol{e} \ : \ \zeta \in \mathbb{R}\right\},$$

where $\boldsymbol{e}$ is a vector of ones.

We first establish the properties of any (auxiliary) AEOs included in $\mathcal{H}$ as well as the induced dynamical system $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$. We use $\leq$ (resp. $\geq$) to denote the elementwise less (resp. greater) than or equal to for the comparison of the vectors.

THEOREM 2 **(Partial Order Preserving, Existence and Uniqueness)**. *Any AEO $\boldsymbol{h}(\cdot) \in \mathcal{H}$ satisfies the following properties (i) and (ii), and the induced dynamical system $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ defined by fixed-point iteration $\boldsymbol{\mu}(t) = \boldsymbol{h}(\boldsymbol{\mu}(t-1))$ satisfies the following property (iii):*

(i) $\boldsymbol{h}(\boldsymbol{a}) \leq \boldsymbol{h}(\boldsymbol{b})$ if $\boldsymbol{a} \leq \boldsymbol{b}$.

(ii) There exists a unique fixed-point solution $\boldsymbol{\mu}^* \in \mathbb{R}^{|V|}$ with $\boldsymbol{h}(\boldsymbol{\mu}^*) = \boldsymbol{\mu}^*$.

(iii) For any $\boldsymbol{\mu}(0) \in \mathbb{R}^{|V|}$, the dynamical system $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ satisfies $\lim_{t\to\infty} \boldsymbol{\mu}(t) = \boldsymbol{\mu}^*$.

The proof of Theorem 2(i) follows from the definition of $\boldsymbol{h}(\cdot)$. Theorem 2(ii) and (iii) are consequent of the fact that $\boldsymbol{h}(\cdot)$ is a contraction mapping. Together, Theorem 2 shows that the dynamical system $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ converges to a unique point regardless of the initial adoptions. These properties of $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ ensure that for any diffusion instance under Assumptions 1 and 2, we can always find a well-defined FPA solution $\boldsymbol{\mu}^*$ for limiting adoption probability $\boldsymbol{q}^*$ by solving the system of equations $\boldsymbol{h}(\boldsymbol{\mu}^*) = \boldsymbol{\mu}^*$.

To show that $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ is close to $\{\boldsymbol{q}(t)\}_{t=0}^{\infty}$, we need to tackle the following two challenges: (i) The adoption states are neither independent in time nor among different agents. When focusing on the adoption utility at each time step, the agent's decision depends on the one-hop neighbors' behavior at the previous time point. When time evolves, the correlation compounds and spreads across the network. For some simple network structures, such as chains, trees and directed acyclic graphs, one may characterize the correlation step by step along the orientation of the edges. However, the problem becomes much more complicated when we assume an arbitrary network structure, as we cannot decompose the network influence by topology. (ii) The CDF function F_ϵ is nonlinear and therefore potentially complex.

However, we show in the next lemma that Assumptions 1 and 2 guarantee that the covariances of any pairs of agents are bounded by the inverse of the minimal indegree $N_{\min}$ at each time step. The intuition behind this is that the network effect of an agent depends on all the in-neighbors. When an agent has more in-neighbors, the network effect originates from more agents, and thus, the correlation between a single pair of agents has the incentive to diminish.

Lemma 1 **(Bounded Covariance)**. *Given a diffusion instance with minimum indegree $N_{\min}$, for any pair of agents $i, i' \in V, i \neq i'$ and $t \geq 0$, the covariance between the adoption states is upper bounded by*

$$Cov(Y_i(t), Y_{i'}(t)) \leq \frac{0.25(L\beta)^2}{1-(L\beta)^2} \frac{1}{N_{\min}}.$$

At a high level, the proof of Lemma 1 decomposes the covariance at each time step and is preceded by bounding each part, step by step. By the law of total covariance, $\mathrm{Cov}(Y_i(t), Y_{i'}(t))$ can be decomposed into two parts conditional on the adoption states in the previous time step as $\mathbb{E}_{\boldsymbol{Y}(t-1)}\left[\mathrm{Cov}(Y_i(t), Y_{i'}(t) | \boldsymbol{Y}(t-1))\right]$ and $\mathrm{Cov}_{\boldsymbol{Y}(t-1)}\left(\mathbb{E}[Y_i(t)|\boldsymbol{Y}(t-1)], \mathbb{E}[Y_{i'}(t)|\boldsymbol{Y}(t-1)]\right)$. The first term equals 0 because the agents' decisions are independent of each other when the adoption states of the previous time step are given. The second term is on the order of $O(1/N_{\min})$ when $t = 1$ and can be iteratively bounded when expressed as Eqn. (2). Under Assumption 2, $\mathrm{Cov}(Y_i(t), Y_{i'}(t))$ increases at a diminishing rate as time t evolves and is consistently on the order of $O(1/N_{\min})$. The covariance upper bound established by Lemma 1 also implies the necessity of Assumption

2, suggesting that the network effect strength β cannot be too large for the diffusion to be well behaved.

With the bounded covariance, we consequently show that the adoption probability trajectory does not deviate far from the dynamical system generated via the FPA scheme. Although $\{\boldsymbol{q}(t)\}_{t=0}^{\infty}$ does not have a closed-form expression, we expect that the transition between consecutive time steps can be bounded by some transformations akin to AEO $\boldsymbol{h}(\cdot)$.

LEMMA 2 **(Bounded Adoption Probability)**. *Given a diffusion instance with minimum indegree $N_{\min}$, for any $t \geq 1$, the adoption probability vector $\boldsymbol{q}(t)$ can be bounded by*

$$\boldsymbol{h}_{-\delta}(\boldsymbol{q}(t-1)) \leq \boldsymbol{q}(t) \leq \boldsymbol{h}_{\delta}(\boldsymbol{q}(t-1))$$

where $\delta = \sqrt{\frac{0.25(L\beta)^2}{1-(L\beta)^2}\frac{1}{N_{\min}}}$.

The proof of Lemma 2 comprises two major steps. The first step is to establish the concentration bound of the local network effect term using the covariance bound from Lemma 1. The second step translates this bound into two auxiliary AEOs in $\mathcal{H}$ such that the difference between $\boldsymbol{q}(t-1)$ and $\boldsymbol{q}(t)$ can be lower and upper bounded by these two AEOs.

Lemma 2 connects the transition of $\{\boldsymbol{q}(t)\}_{t=0}^{\infty}$ to that of the dynamical system $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ and provides a one-step guarantee. Based on this fact, we use a "fixed-point sandwich" technique to prove the final results. We first define a lower bound dynamical system $\{\underline{\boldsymbol{\mu}}(t)\}_{t=0}^{\infty}$ and an upper bound dynamical system $\{\overline{\boldsymbol{\mu}}(t)\}_{t=0}^{\infty}$ as follows.

$$\underline{\mu}_i(t) = \begin{cases} q_i(0) & t = 0 \\ 1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} \underline{\mu}_j(t-1)}{n_i}\right) - \delta & t > 0 \end{cases}, \text{ for all } i \in V$$

$$\overline{\mu}_i(t) = \begin{cases} q_i(0) & t = 0 \\ 1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} \overline{\mu}_j(t-1)}{n_i}\right) + \delta & t > 0 \end{cases}, \text{ for all } i \in V$$

These two systems can also be written as two FP iterations $\underline{\boldsymbol{\mu}}(t) = h_{-\delta}(\underline{\boldsymbol{\mu}}(t-1))$ and $\overline{\boldsymbol{\mu}}(t) = h_{\delta}(\overline{\boldsymbol{\mu}}(t-1))$ using the auxiliary AEOs. Let $\underline{\boldsymbol{\mu}}^*$ and $\overline{\boldsymbol{\mu}}^*$ be the FP solutions to these two systems, respectively. To prove Theorem 1, we demonstrate that the entire trajectories of $\{\boldsymbol{q}(t)\}_{t=0}^{\infty}$ and $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ are close to each other. When time tends to infinity, FP solution $\boldsymbol{\mu}^*$ serves as a good approximation for limiting adoption probability $\boldsymbol{q}^*$. More specifically, the proof of Theorem 1 uses two fixed-point iterations $\{\underline{\boldsymbol{\mu}}(t)\}_{t=0}^{\infty}$ and $\{\overline{\boldsymbol{\mu}}(t)\}_{t=0}^{\infty}$ to bound $\{\boldsymbol{q}(t)\}_{t=0}^{\infty}$. As a result, when t tends to infinity, the maximal deviation $\|\boldsymbol{q}^* - \boldsymbol{\mu}^*\|_\infty$ can be bounded by the distance between the lower and upper FP solutions $\|\underline{\boldsymbol{\mu}}^* - \overline{\boldsymbol{\mu}}^*\|_\infty$.

We remark here that all the aforementioned upper bounds are not applicable to the networks where $N_{\min} = 0$. The reason is that, for agents who do not have in-neighbors, we set the local network effect term in Eqn. 1 to 0. In other words, their decisions are not influenced by others and hence are independent throughout the time horizon. As a result, we can easily analyze these standalone nodes separately while leaving the remaining network structure to the FPA scheme.

4.4. Numerical Experiments on the FPA Solution

In this section, we conduct simulation studies to further validate our FPA scheme under different scenarios. Through extensive numerical experiments, we highlight that, for a wide range of reasonable problem instances, the actual performance of our approximation scheme is *much better* than the theoretical guarantee proved in Theorem 1. In particular, the results show that even for the diffusion instances where the network is small (i.e., $|V|$ is small) and/or sparse (i.e., $|\mathcal{N}_i|$ is small for some agents), the average percentage error of the FPA solution is quite small.

Before presenting our experiments, we remark that there is a fundamental challenge in measuring the accuracy of the FPA solution. For a particular diffusion instance, the approximation accuracy is difficult to measure, as the ground-truth limiting adoption probabilities are unknown beforehand. To obtain the ground truth, as discussed in Section 3, we need to solve the stationary distribution of a large-scale MC, which is intractable in general.[1] Hence, we resort to the MCMC simulation technique to examine the performance of our FPA scheme for random graphs (Section 4.4.1) and real-world networks (Section 4.4.2). In all our experiments, the FP solution $\boldsymbol{\mu}^*$ is calculated via an FP iteration with precision 10^{-5}.

4.4.1. Random Graphs. In this subsection, we test our FPA scheme with Erdős–Rényi random graphs. As we cannot obtain the ground truth for general diffusion instances, hereafter, we use MCMC to estimate the limiting adoption probability. The simulation starts when all the agents in the network are not adopted. The first 1,000 time steps of the simulation are considered the warm-up period and are discarded as being transient. This is sufficient to avoid bias in our study. We include more details about the warm-up period in Appendix B.4. For following Eqn. (4), we set the run length for each simulation replication to be 100,000 steps beyond the warm-up periods and take the adoption frequency of each agent throughout this period as the ground truth of the limiting adoption probability.

We denote a random graph as $G(N,\theta)$ in which there are N nodes, and every possible edge among the nodes occurs independently with probability θ. To show the overall performance of the FPA scheme in the network, we measure the accuracy with the mean absolute percentage error (MAPE) of all the agents in this instance:

$$\text{MAPE} = \frac{1}{N}\sum_{i\in V}\frac{|\mu_i^* - q_i^*|}{q_i^*}\cdot 100\%$$

[1] For some highly structured symmetric networks (such as star networks and complete networks), solving the stationary distribution is feasible. See Appendix B.3 for details.

For the experiments in this subsection, we assume intrinsic value $v_i \overset{i.i.d.}{\sim} \mathcal{U}(-4,0)$ and random noise $\epsilon_i(t) \overset{i.i.d.}{\sim} \text{Logistic}(0,1)$[2].

In the following, we conduct three experiments to understand the performance of the FPA scheme from different perspectives: (i) the sensitivity with regard to the network effect strength β, (ii) the accuracy with regard to the network structure (i.e., network size, density, and indegree), and (iii) the efficiency with regard to the network size.

The sensitivity with regard to β. We first test the sensitivity of the FPA solution's accuracy with respect to the network effect strength β on a small network. When β is close to 0, it is not surprising that the FPA scheme can have a superior performance. When β increases, the network effect plays a larger role in the limiting adoption probability, and the approximation quality degrades in the meantime. According to Assumption 2, we restrict the network effect parameter to be $\beta < \frac{1}{L} = 4$ and test with β from 0 to 3.5. For each β, we generate 50 different diffusion instances with random graph $G(50, 0.1)$. To show the effectiveness of the approximation, we benchmark on the misspecified model where the network diffusion is not incorporated. This benchmark is the standard binary choice model, where the adoption utility only consists of an intrinsic value and random noise. In Figure 2, we show the MAPE of both models. We observe that the MAPE of the FPA solution

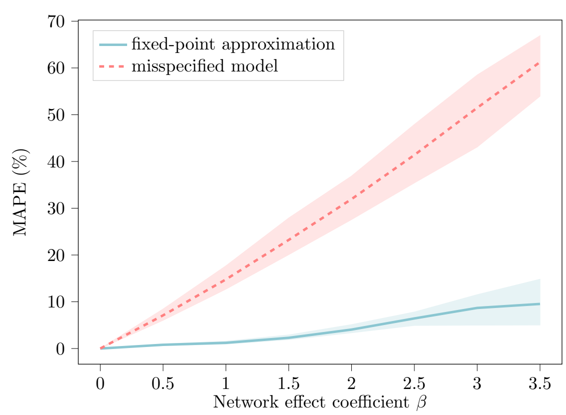


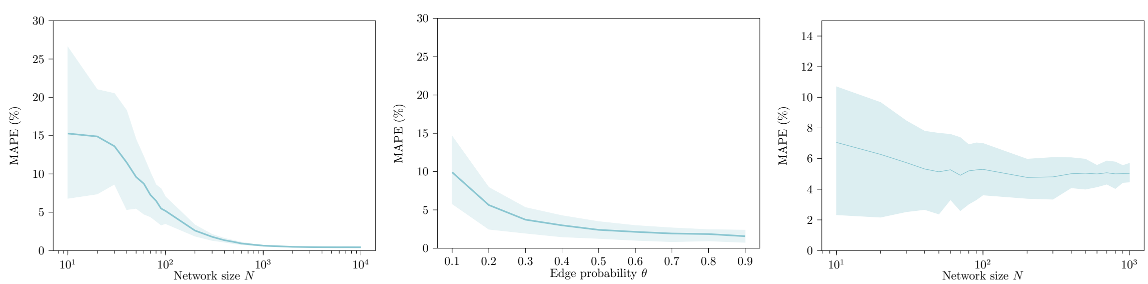
Figure 2 MAPE of the FPA solution versus network effect parameter (Color shaded area denotes the 95% confidence interval)

increases with β but it increases at a much slower rate when compared with the misspecified model. In particular, when $\beta = 3.5$, the network diffusion effect becomes dominant in the adoption utility. The MAPE of the FPA solution is approximately 9.53%, which is less than one-sixth of the misspecified model (i.e., 61.25%). This shows that even with such a small network and strong network effects, our method performs well. Hence, we fix a relatively large network effect strength $\beta = 3.5$ in the subsequent experiments to offer insights into the worst-case scenario.

[2] The CDF of Logistic(μ, σ) is given by $1/(1 + \exp\{-(x-\mu)/\sigma\})$.

<u>*The accuracy with regard to network structure.*</u> The parameters of a random graph, N and θ, naturally represent the network size and density, respectively. We vary these two parameters to generate structurally different networks. For each network structure, we test our method over 50 diffusion instances.

Figure 3a illustrates the performance of the FPA scheme for different network sizes. We fix $\theta = 0.1$ and select N from 10 to 10,000. We can observe that the MAPE decreases rapidly when the network becomes larger. For diffusion instances with more than 1,000 nodes, the average MAPE is less than 0.7%. Figure 3b illustrates the performance of the FPA scheme over different network densities. We fix $N = 50$ and change θ from 0.1 to 0.9. We also notice a decreasing trend when the network becomes dense. Figure 3c shows the performance of the FPA scheme when the expected number of neighbors remains the same. We generate a sequence of networks where we fix $N\theta = 10$ and select N from 10 to 1,000. In this case, the average indegree of all diffusion instances is only 10. Overall, the approximation error is small, with an average MAPE of less than 8%. Furthermore, as the network size increases, the MAPE also slightly decreases. In conclusion, the experimental results demonstrate that the FPA scheme can provide an accurate approximation for arbitrary diffusion instances, especially for large and dense networks, complementing our theoretical analysis.



(a) Network size (b) Network density (c) Fixed indegree

Figure 3 MAPE of the FPA solution versus network structures. Left: Network size (x axis is in log scale); Right: Network density; Below: Fixed indegree (x axis is in log scale). (Color shaded area denotes the 95% confidence interval)

<u>*The efficiency with regard to network size.*</u> We compare the CPU time between the calculation of $\boldsymbol{p}^*$ and $\boldsymbol{\mu}^*$. To make the results comparable, we terminate the MCMC procedure when its real-time MAPE first falls below that of the FPA solution. As the runtimes of both the FPA scheme and MCMC are closely related to the network size, we fix $\theta = 0.1$ and choose N from 10 to 10,000. Figure 4 shows that when the network becomes large, the runtime for MCMC increases dramatically. The runtime for the FPA scheme also increases but remains at a low value consistently. For a

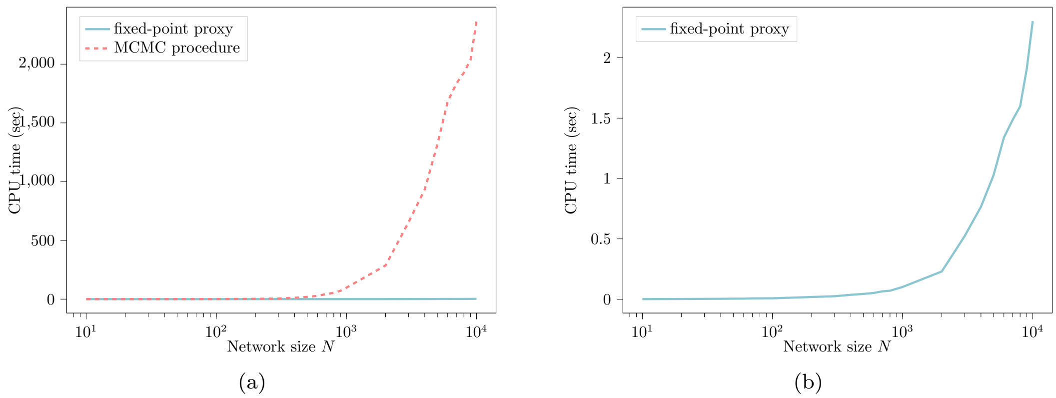


Figure 4 Execution time of FPA scheme and MCMC method (x axis is in log scale)

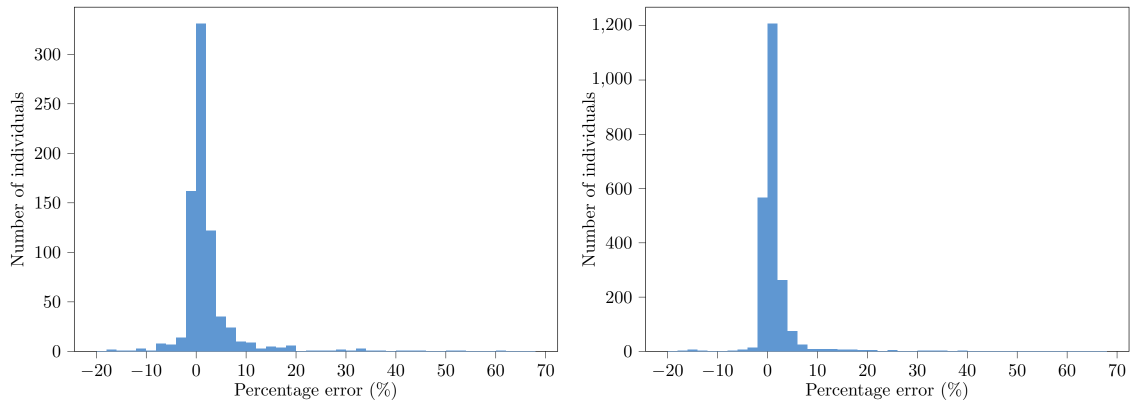
network with 10,000 nodes, approximately 40 minutes are required for MCMC to acquire a similar performance as the FPA scheme, which finishes in 2.3 seconds.

In conclusion, without compromising accuracy, the FPA scheme enjoys great advantages in computational efficiency compared with MCMC for all the diffusion instances examined. Therefore, it is promising to use the FPA scheme to effectively characterize the diffusion process for a large variety of social networks.

4.4.2. Real-world Networks. Our numerical experiments in Section 4.4.1 focus on the random graph setting. Networks in this setting have been criticized for failing to capture certain real-world phenomena. For example, they assume the degrees are almost uncorrelated between neighbors and overlook the possibility that high-degree agents tend to have high-degree neighbors, a natural outcome of homophily (Jackson 2010). As such, to further demonstrate the superior performance of the FPA scheme in more practical settings, we test it on real-world network instances from the Network Repository Rossi and Ahmed (2015). This repository includes thousands of networks in more than 30 domains, from biological to social network data. We choose two Facebook networks, which are social friendship networks extracted from Facebook consisting of people (nodes) with the edges representing friendship ties. For each diffusion instance, we still set $\beta = 3.5$, the intrinsic value $v_i \overset{i.i.d.}{\sim} \mathcal{U}(-4, 0)$, and random noise $\epsilon_i(t) \overset{i.i.d.}{\sim} \text{Logistic}(0, 1)$, as in the previous discussion. We provide an overview of these two networks as follows. The raw data as well as more data statistics can be found on the website of the Network Repository.

Network instances:

- *Caltech36*: The instance includes 765 agents with an average number of neighbors of 43, and the maximum (resp. minimum) degree is 246 (resp. 1).
- *Amherst41*: The instance includes 2,235 agents with an average number of neighbors of 81, and the maximum (resp. minimum) degree is 467 (resp. 1).

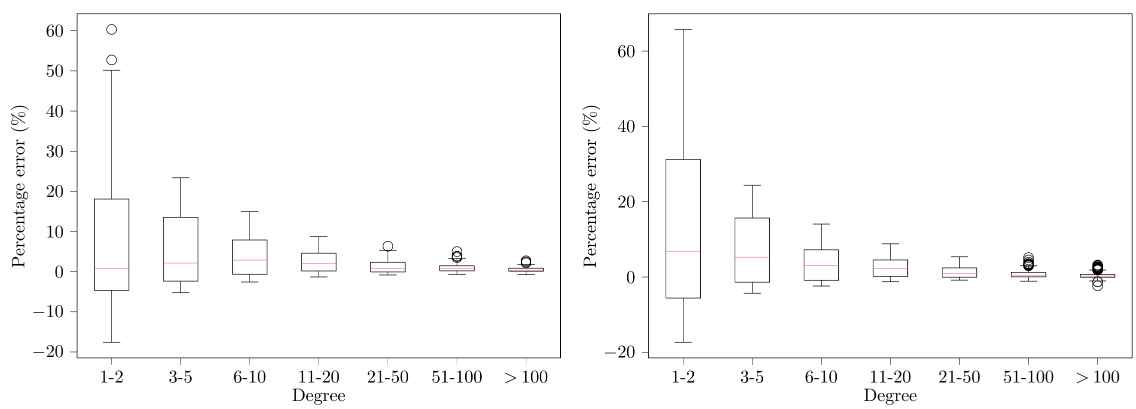


(a) *Caltech36* (b) *Amherst41*

Figure 5 Distribution of percentage error of fixed-point proxy in a single network instance

Figure 5 plots the percentage error among all the agents in each diffusion instance. In general, we observe that the errors concentrate at approximately 0. The instance *Caltech36* has an MAPE of 3.10% while the instance *Amherst41* has an MAPE of 1.79%.

We still observe that very few nodes in the network have relatively large absolute percentage errors. We demonstrate in Figure 6 that these nodes usually have small degrees. It is important



(a) *Caltech36* (b) *Amherst41*

Figure 6 Range of percentage errors for agents with similar degrees in a single network instance

to note that the range of errors shrinks rapidly when the degree increases. For example, in such practical instances, for nodes having more than ten neighbors, the percentage error of the FPA solution is usually smaller than 10%. In reality, agents with more neighbors usually play a much more important role from the platform's perspective. Therefore, we also measure the performance using weighted MAPE (WMAPE), where the error of each agent is weighted by degree as

$$\mathrm{WMAPE} = \frac{1}{\sum_{i\in V} n_i} \cdot \sum_{i\in V}\left(n_i \cdot \frac{|q_i^* - \mu_i^*|}{q_i^*}\right) \cdot 100\%.$$

The WMAPE for instance *Caltech36* and *Amherst41* is 1.20% and 0.75%, respectively. We notice that the WMAPE is much less than the MAPE measured before, while the larger instance consistently exhibits a better result. These two experiments show that the FPA scheme provides an excellent approximation for the diffusion model in real-world network instances.

For the completeness of the experiments on real-world networks, we repeat the experiments 10 times and include the statistics of the results (e.g., MAPE, WMAPE and the percentage error by degree) in Appendix B.5. The average over 10 repetitions is close to the results we showed before.

5. Applications of the Fixed-point Approximation Scheme

The FPA scheme can be applied to many classical operational decision problems that involve network diffusion. In this section, we illustrate this with two important examples. One is the influence maximization (IM) problem (Kempe et al. 2003, Li et al. 2018) in network analysis, and the other is the pricing problem with network effects (Candogan et al. 2012, Du et al. 2016) in revenue management. Hereafter, we assume that the network structure and intrinsic values are known to the platform and confine our analysis of the optimization problems to a given diffusion instance.

5.1. Influence Maximization Problem

In the IM problem, we select up to a set of K seed users to adopt the service at the beginning of the planning horizon to maximize the expected adoptions in the long run. For example, the service provider may select a few social network influencers to promote the service to the general audience. We additionally require that the adoptions of the seed users be irreversible in the IM problem in the nonprogressive diffusion setting, as opposed to only changing their initial states. The reason is twofold. First, from an application perspective, this essentially assumes that the influence of the seed users is long-lasting, as often is the case in practice. Second, as demonstrated in Proposition 1 and its proof, Assumptions 1 and 2 guarantee a unique aperiodic recurrent class of states, so under the same network parameters, the initial state change does not alter the long-run limit. Instead, by requiring the seed users to remain adopted throughout the entire time horizon, we effectively change the problem parameter. One can either interpret this as increasing the intrinsic values of the seed users to be large enough so they almost always adopt, or increasing the intrinsic values of the seed user neighbors so they are always subject to the positive externalities of the seed users. We can formulate the IM problem using the FPA scheme as a special case of Eqn. (7) and define $R(S)$ as follows,

$$\underset{\boldsymbol{\mu}, S \subseteq V}{\text{maximize}} \quad \sum_{i \in V} \mu_i \tag{9a}$$

$$\text{subject to} \quad \mu_i = 1, \qquad \forall i \in S, \tag{9b}$$

$$\mu_i = 1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i}\mu_j}{n_i}\right),\ \forall i \in V\backslash S. \tag{9c}$$

The equality constraints imply that when given the seed set S, the value of $\boldsymbol{\mu}$ is also uniquely determined. Hence, we define $R(S)$ as the influence function when all the agents in S are set to be permanently adopted. In this case, it is straightforward to adapt the proof of Theorem 1 to show that this FPA-based formulation still approximates the limit of diffusion well. As the IM problem is known to be NP-hard, most related studies developed algorithms based on the submodularity of the influence function. We claim that given some assumptions on the intrinsic value and CDF F_ϵ, $R(S)$ is also a submodular set function.

ASSUMPTION 3 **(Convexity on Part of the CDF)**. *The CDF F_ϵ is convex on $\mathcal{D}$, where $\mathcal{D}$ is the support of the intrinsic value v_i for all $i \in V$.*

The assumption for convexity seems strong at first glance. However, we point out that it still covers a wide range of commonly studied use cases. For example, the classical setting of the IM problem is a special case of our model under this assumption. Recall that we can recover the classical linear threshold model by setting $v_i = -0.5$ and $\epsilon_i(t) \sim \mathcal{U}(-0.5, 0.5)$ for all $i \in V$ and $t \geq 0$. Therefore, on the interval $(-\infty, 0.5]$, CDF F_ϵ can be written as $F_\epsilon(x) = \mathbb{1}\{x \geq -0.5\}\cdot(x+0.5)$, which is convex. Some other diffusion instances that are related to common utility models can also satisfy Assumption 3: we list some examples: (i) Linear probability model: $v_i \geq -c$ and $\epsilon_i(t) \sim \mathcal{U}(-c, c)$ for all $i \in V$, $t \geq 0$. (ii) Logit model: $v_i \geq 0$ and $\epsilon_i(t) \sim \text{Logistic}(0, \sigma)$ and for all $i \in V$, $t \geq 0$. (iii) Probit model: $v_i \geq 0$ and $\epsilon_i(t) \sim \mathcal{N}(0, \sigma)$ and for all $i \in V$, $t \geq 0$. For many general distributions, the convexity assumption essentially requires the intrinsic values to be appropriately lower bounded. With Assumption 3, the following theorem shows the submodularity of $R(S)$.

THEOREM 3 **(Submodularity of total influence function)**. *Under Assumption 3, the total influence function $R(S)$ is a submodular set function with regard to the seed set S.*

Therefore, we can apply the well-known greedy algorithm (Nemhauser et al. 1978) to solve the IM problem in our setting. For diffusion instances that conform with Assumption 3, the greedy algorithm can provide an $(1 - 1/e)$-approximation solution. This discussion reveals the benefits of formulating an IM problem with the FPA scheme. The FPA scheme preserves the submodularity of the influence function, which is at the core of the IM algorithms. Additionally, from a computational perspective, the greedy algorithm needs to call the oracle that calculates the limiting expected adoptions $\mathcal{O}(NK)$ times (Nemhauser et al. 1978). The fast computation of the FPA solution can improve the execution time compared with simulation techniques in a nontrivial way.

5.1.1. Experiments on IM Problem. In the experiments, we consider two scenarios, one satisfies Assumption 3 and thus leads to a submodular influence function, while the other does not. For both scenarios, we assume that the intrinsic value $v_i \overset{i.i.d.}{\sim} \mathcal{U}(-4,0)$ and $\beta = 3.5$. In addition, we assume the random noise to be $\epsilon_i(t) \overset{i.i.d.}{\sim} \mathcal{U}(-4,4)$ in the submodular case, while $\epsilon_i(t) \overset{i.i.d.}{\sim}$ Logistic$(0,1)$ in the nonsubmodular case.

The well-known greedy framework selects one user at each iteration that leads to the largest total adoptions. We refer to the algorithm that embeds the FPA solution into this greedy framework for the total influence evaluation as the greedy-FP algorithm. We randomly generate some small network instances to illustrate that greedy-FP can find a near-optimal solution. Although there is no theoretical guarantee for the nonsubmodular case, it is interesting to observe from the results that the FPA solutions are still of good quality. For either scenario, we generate 100 diffusion instances with random graph $G(15,0.5)$ and set the number of seed users to be 5. We enumerate all the subsets to find the optimal seed set and evaluate the diffusion influence using MCMC. In Table 1, we show the numerical results of the greedy-FP algorithm.

Table 1 Numerical results of greedy-FP algorithm for IM problem

Scenario	Percentage of instances where the optimal seeding is recovered	Optimality Gap (%) Mean	Max
Submodular	91	0.0194	0.4704
Non-submodular	84	0.0685	1.7444

We notice that in both the submodular and nonsubmodular scenarios, the greedy-FP algorithm can generate a near-optimal IM solution and even uncover the exact optimal solution for a large portion of instances. Meanwhile, the greedy-FP algorithm has a slightly better performance in the submodular case than in the nonsubmodular case but even in the nonsubmodular problem instances, it remains quite practical.

Furthermore, we choose a real-world network—*Caltech36* as introduced in Section 4.4.2 and compare the performance of greedy-FP with the traditional IM heuristics. Recall that the instance includes 765 agents with an average number of neighbors of 43. We define several benchmark strategies as follows. The DEG and EIG schemes are motivated by the important role of the centrality measures in diffusion discussed in the network economics literature (e.g., Ballester et al. 2006, Jackson 2010). We include them for completeness, but as substantiated in the numerical experiments, by overlooking the idiosyncratic features of the agents, these schemes are dominated by the FPA-based heuristic.

Benchmarks:

- *Greedy and MCMC* (greedy-MCMC): This is the classical algorithm used for the IM problem. The MCMC is embedded into a greedy framework for influence evaluation. The length of the MCMC run is set to 100,000 after the warm-up period.
- *Greedy and the FPA solution* (greedy-FP): This is our proposed algorithm. We embed the FPA solution into a greedy framework for an influence evaluation.
- *Greedy and low resolution MCMC* (greedy-l-MCMC): The MCMC is embedded into a greedy framework for an influence evaluation. The length of the MCMC run is set to 50 so that the runtime is at the same scale as that of the FPA scheme.
- *Degree centrality* (DEG): Set K users with the largest degree to be seed users.
- *Eigenvector centrality* (EIG): Set K users with the largest eigenvector centrality (Bonacich 1972) to be seed users.
- *Model misspecification without network effect* (MM): This benchmark considers the misspecified model that ignores the network effect in the IM problem. This is the same as setting K users with the smallest intrinsic value to be seed users.
- *Random* (RAN): Randomly select K users to be seed users.

Figure 7 demonstrates the relative loss of the expected limiting adoptions compared with greedy-MCMC against the number of seed users. Similarly, we also consider both the submodular and nonsubmodular cases.

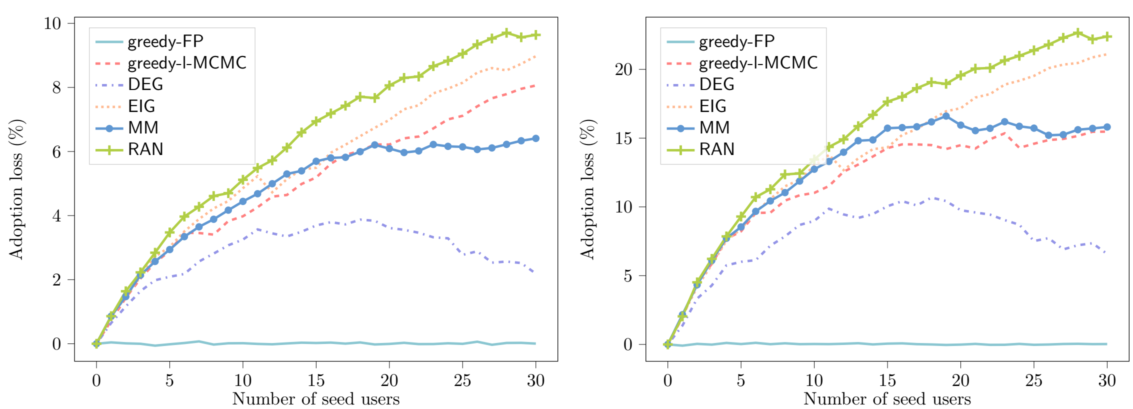


Figure 7 Performance of different IM algorithms. Left: submodular case; Right: non-submodular case.

When the number of seed users increases from 0 to 30, the difficulty of the IM problem increases since the number of feasible solutions also increases. We observe that regardless of the number of seed users, the performance of greedy-FP matches that of greedy-MCMC nearly perfectly. It significantly outperforms all the other benchmarks. This is no surprise to us; this is driven by the high accuracy of the FPA scheme. In particular, we also notice that the performance of the greedy framework with the MCMC method degrades drastically when the simulation length of the MCMC procedure is small. Compared with greedy-I-MCMC, greedy-FP achieves an improvement of 8.90% and 18.42% when $K=30$ in the submodular and nonsubmodular cases. In short, we conclude that, by offering a significant efficiency gain, greedy-FP outperforms greedy-MCMC in solving the IM problem.

5.2. Pricing Problem on a Social Network

Network effects can play an important role in determining customers' preference for a service or product item. In view of this, an emerging literature takes into account the network effects or diffusion effects in the revenue management problem (Du et al. 2016, 2018, Wang and Wang 2017, Chen and Shi 2022, Chen and Chen 2021, Gopalakrishnan et al. 2022). As we discuss next, our formulation of the pricing problem with the diffusion effects incorporated is naturally connected to the existing ones in this literature. The formulations borne out of these works are usually axiomatic or game-theoretic but our previous discussion on the FPA scheme shows that such formulations are also justified in a dynamic, stochastic setting.

To investigate the impact of prices on adoption utility, we separate the charged price term from the intrinsic utility. Suppose that the price offered to user i is p_i. The adoption utility at time t can then be written as

$$u_i(t) = v_i - \alpha p_i + \beta \cdot \frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{n_i} + \epsilon_i(t)$$

where α represents the price sensitivity. In other words, the firm uses price operational levers to steer consumers' adoption decisions. To formalize the pricing problem in a general setting, we allow the offering of different prices to different consumers. Indeed, some platforms have the power to implement price discrimination against customers. Suppose the platform allows at most M different prices; we use $\boldsymbol{p}\in\mathbb{R}^M$ to denote the price vector. Meanwhile, we define a transformation matrix $W\in\mathbb{R}^{N\times M}$, where $W_{ik}=1$ if consumer i is assigned with the k-th price; otherwise, $W_i=0$. When $M=N$ and $W=I_N$, customers face idiosyncratic prices. When $M=1$ and $W=\boldsymbol{e}_N$, customers face a homogeneous price. We assume that the transformation matrix W is known by the platform; that is, the platforms have already decided to offer different prices to M different customer types. In the pricing problem, we are interested in finding an optimal price vector for profit maximization.

With the help of the FPA scheme, we can explicitly formulate the pricing problem as follows. Although in principle $\boldsymbol{p}$ is the only decision variable, we find it convenient to introduce the long-run (approximate) adoption probability $\boldsymbol{\mu}$ as an additional decision variable and use the constraint (10b) to link it with $\boldsymbol{p}$.

$$\underset{\boldsymbol{\mu},\boldsymbol{p}}{\text{maximize}} \quad \boldsymbol{\mu}^\top W\boldsymbol{p} \tag{10a}$$

$$\text{subject to} \quad \mu_i = 1 - F_\epsilon\left(-v_i + \alpha \sum_{k=1}^{M} W_{ik} p_k - \beta \frac{\sum_{j\in\mathcal{N}_i} \mu_j}{n_i}\right), \ \forall i \in V. \tag{10b}$$

Some remarks are in order. First, different from the literature, the pricing problem (10) aims at approximately optimizing the total long-run total adoptions in a dynamic, stochastic environment with a general formulation that incorporates various practical considerations such as network topology, local network effects and customer heterogeneity. Theorem 1 guarantees that its solution offers a near-optimal solution for practical diffusion problems, as also evidenced by the subsequent numerical experiments. Second, note that (10) is in general nonconvex due to the constraint (10b). Additionally, the distribution of random noise ϵ also impacts the exact formulation as well as the difficulty of the problem. The CDFs of some common utility models, such as the probit models, are complex and may not even possess a closed-form expression. Hereafter, we emphasize the case with the logit model where $\epsilon_i(t) \overset{i.i.d.}{\sim} \text{Logistic}(0,1)$ $(F_\epsilon(x) = 1/(1+e^{-x}))$. Then, the formulation naturally relates to the existing revenue management literature (Li and Huh 2011, Gallego and Wang 2014, Golrezaei et al. 2020, Chen and Shi 2022), in which a proven useful technique to analyze such a pricing problem is to transform it into an optimization problem in which the market share is used as the primary decision variable. Motivated by this, we consider the problem in both the adoption probability and the price space.

5.2.1. Profit Maximization in the Adoption Probability Space. When considering the adoption probability space, the pricing problem becomes less demanding when certain technical conditions hold. In particular, we assume a standard logit model for the random noise term. Furthermore, in a perfect price discrimination environment $(M = N,\ W = I_N)$ where the platform can provide an idiosyncratic price/subsidy to each consumer and there are no price constraints, one can reformulate the problem as follows.[3]

$$\underset{\boldsymbol{\mu},\boldsymbol{p}}{\text{maximize}} \quad \boldsymbol{\mu}^\top \boldsymbol{p} \tag{11a}$$

$$\text{subject to} \quad \mu_i = 1 - \frac{1}{1+\exp\{v_i + \beta \sum_{j\in\mathcal{N}_i} \mu_j/n_i - \alpha p_i\}}, \ \forall i \in V. \tag{11b}$$

By rearranging terms, we can solve for the price from (11b) as

$$p_i(\boldsymbol{\mu}) = \frac{1}{\alpha}\left(v_i + \beta \sum_{j\in\mathcal{N}_i} \frac{\mu_j}{n_i} + \ln\frac{1-\mu_i}{\mu_i}\right), \ \forall i \in V \tag{12}$$

[3] Possible negative prices mean that the platform can subsidize some users, in particular those who might have a large influence on the network. The platform incurs losses for these customers to promote a larger overall profit, as commonly found in practice.

The optimization problem can then be reformulated in the adoption probability space as follows.

$$\underset{\boldsymbol{\mu}}{\text{maximize}} \quad \sum_{i\in V}\frac{1}{\alpha}\left(v_i+\beta\sum_{j\in\mathcal{N}_i}\frac{\mu_j}{n_i}+\ln\frac{1-\mu_i}{\mu_i}\right)\mu_i \tag{13a}$$

$$\text{subject to} \quad 0\le\mu_i\le 1,\ \forall i\in V. \tag{13b}$$

Note that local network effects are captured in the term $\sum_{j\in\mathcal{N}_i}\mu_j/n_i$ for each customer $i\in V$. When $\beta=0$, this term does not play a role; the problem can then be separated into the sum of relative entropy functions and therefore is concave. This property is generally preserved when β is small enough.

Theorem 4 **(Concavity of Price Optimization)**. *The objective of the pricing problem* (13) *is concave in* $\boldsymbol{\mu}$ *if and only if* $0<\beta\le 3.375$.

Theorem 4 states that when the network diffusion parameter satisfies $0<\beta\le 3.375$, problem (13) is a convex optimization problem and the optimal adoption probability $\boldsymbol{\mu}^*$ can be solved by standard optimization techniques (i.e.. gradient methods). Given $\boldsymbol{\mu}^*$, we can recover the optimal prices by Eqn. (12). Furthermore, we notice that both Theorem 4 and Assumption 2 require the network effect parameter to be relatively small, while Theorem 4 seems more restrictive.

5.2.2. Profit Maximization in the Price Space. In more general settings, we need to study profit maximization in the price space. Particularly, we represent the adoption probability as an implicit function of price, and write the objective function as a univariate function with regard to $\boldsymbol{p}$. We write the adoption probability $\boldsymbol{\mu}(\boldsymbol{p})$ as a function of price to highlight that it is determined by the prices. According to Theorem 2, there exists a one-to-one correspondence from $\boldsymbol{p}$ to $\boldsymbol{\mu}(\boldsymbol{p})$. Therefore, we can derive the gradient of the profit function $\Pi(\boldsymbol{p})$ as follows.

$$\frac{d\Pi(\boldsymbol{p})}{d\boldsymbol{p}}=\frac{d\boldsymbol{\mu}(\boldsymbol{p})}{d\boldsymbol{p}}\cdot W\cdot\boldsymbol{p}+W^\top\cdot\boldsymbol{\mu}(\boldsymbol{p}). \tag{14}$$

In Eqn. (14), we note that the gradient of $\boldsymbol{\mu}(\boldsymbol{p})$ is not explicitly given. To simplify the problem, we notice that by definition it holds that $\boldsymbol{\mu}(\boldsymbol{p})=\boldsymbol{h}(\boldsymbol{p},\boldsymbol{\mu}(\boldsymbol{p}))$. We take derivatives on both sides to obtain

$$\frac{d\boldsymbol{\mu}(\boldsymbol{p})}{d\boldsymbol{p}}=\frac{\partial\boldsymbol{h}(\boldsymbol{p},\boldsymbol{\mu}(\boldsymbol{p}))}{\partial\boldsymbol{p}}+\frac{d\boldsymbol{\mu}(\boldsymbol{p})}{d\boldsymbol{p}}\cdot\frac{\partial\boldsymbol{h}(\boldsymbol{p},\boldsymbol{\mu}(\boldsymbol{p}))}{\partial\boldsymbol{\mu}(\boldsymbol{p})}.$$

By rearranging the terms, we obtain

$$\frac{d\boldsymbol{\mu}(\boldsymbol{p})}{d\boldsymbol{p}}\cdot\left(I-\frac{\partial\boldsymbol{h}(\boldsymbol{p},\boldsymbol{\mu}(\boldsymbol{p}))}{\partial\boldsymbol{\mu}(\boldsymbol{p})}\right)=\frac{\partial\boldsymbol{h}(\boldsymbol{p},\boldsymbol{\mu}(\boldsymbol{p}))}{\partial\boldsymbol{p}}.$$

Matrix $(I - \partial \boldsymbol{h}(\boldsymbol{p},\boldsymbol{\mu}(\boldsymbol{p}))/\partial \boldsymbol{\mu}(\boldsymbol{p}))$ is guaranteed to be invertible. The reason is that, by Theorem 2, we know that $\boldsymbol{h}$ is a contraction mapping and $\|\partial \boldsymbol{h}(\boldsymbol{p},\boldsymbol{\mu}(\boldsymbol{p}))/\partial \boldsymbol{\mu}(\boldsymbol{p})\|_\infty < 1$. Therefore, the gradient of the profit function can be written as

$$\frac{d\Pi(\boldsymbol{p})}{d\boldsymbol{p}} = \frac{\partial \boldsymbol{h}(\boldsymbol{p},\boldsymbol{\mu}(\boldsymbol{p}))}{\partial \boldsymbol{p}} \cdot \left(I - \frac{\partial \boldsymbol{h}(\boldsymbol{p},\boldsymbol{\mu}(\boldsymbol{p}))}{\partial \boldsymbol{\mu}(\boldsymbol{p})}\right)^{-1} \cdot W \cdot \boldsymbol{p} + W^\top \cdot \boldsymbol{\mu}(\boldsymbol{p}) \tag{15}$$

As a result, when the price vector $\boldsymbol{p}$ is given, we can directly derive the gradient of the profit function. Consequently, we can apply standard gradient descent techniques for nonconvex optimization to achieve a near-optimal solution.

With an eye toward implementation, we also notice that (15) involves the derivation of the gradient, which requires computing the inverse of an $|V| \times |V|$ matrix. When the network is large and dense, this calculation becomes intimidating. However, we notice that $\|\partial \boldsymbol{h}(\boldsymbol{p},\boldsymbol{\mu}(\boldsymbol{p}))/\partial \boldsymbol{\mu}(\boldsymbol{p})\|_\infty < 1$, and therefore, the spectral radius of $\partial \boldsymbol{h}(\boldsymbol{p},\boldsymbol{\mu}(\boldsymbol{p}))/\partial \boldsymbol{\mu}(\boldsymbol{p})$ is smaller than 1, so we can expand the inverse as (e.g., see Hogben 2006),

$$\left(I - \frac{\partial \boldsymbol{h}(\boldsymbol{p},\boldsymbol{\mu}(\boldsymbol{p}))}{\partial \boldsymbol{\mu}(\boldsymbol{p})}\right)^{-1} = \lim_{n\to\infty} \sum_{\ell=0}^{k} \left(\frac{\partial \boldsymbol{h}(\boldsymbol{p},\boldsymbol{\mu}(\boldsymbol{p}))}{\partial \boldsymbol{\mu}(\boldsymbol{p})}\right)^{\ell}.$$

This leads to the following k-th order approximate gradient:

$$\frac{d\Pi(\boldsymbol{p})}{d\boldsymbol{p}} \approx \widetilde{G}_k(\boldsymbol{p}) = \frac{\partial \boldsymbol{h}(\boldsymbol{p},\boldsymbol{\mu}(\boldsymbol{p}))}{\partial \boldsymbol{p}} \cdot \left(I + \sum_{\ell=1}^{k} \left(\frac{\partial \boldsymbol{h}(\boldsymbol{p},\boldsymbol{\mu}(\boldsymbol{p}))}{\partial \boldsymbol{\mu}(\boldsymbol{p})}\right)^{\ell}\right) \cdot W \cdot \boldsymbol{p} + W^\top \cdot \boldsymbol{\mu}(\boldsymbol{p}) \tag{16}$$

for the pricing problem. We expect such an easy-to-compute approximate gradient to lead to a significant efficiency gain, as is usually the case in the literature regarding approximate gradient descent (Ruder 2016). Previous works have applied similar low-order approximations for network effects for different purposes (e.g., see Candogan et al. 2012, Zeng et al. 2021). In subsequent numerical experiments, we find that $k = 2$ works very well in practice, leading to near-optimal solutions very quickly. As the final remark here, we mention that this discussion in the price space is valid under any noise distributions, and that the gradient-based approach in the price space can be applied to more sophisticated cases with additional constraints on prices (e.g., box constraints).

5.2.3. Experiments on Pricing Problems. In the experiments for pricing problems on a social network, the first issue is that the optimal pricing problem under the original diffusion model seems impossible to derive. We test over different randomly generated instances and find that the profits calculated via MCMC and the FPA scheme are quite close, with a percentage error almost uniformly bounded by 0.5% in our experiment. We show the details in Appendix C.2. Hereafter, we compare the pricing scheme under the FPA scheme as the default. We assume $\epsilon_i(t) \overset{i.i.d.}{\sim} \text{Logistic}(0,1)$, which follows the theoretical analysis in Section 5.2. We study two extreme scenarios, the *perfect price discrimination* case, where each consumer is offered a personal price, and the *public price* case, where all consumers receive the same price.

In the perfect price discrimination scenario, we test three different algorithms. The first algorithm is the gradient descent method in the adoption probability space (grad-PROB). With a network effect parameter that satisfies Theorem 4, grad-PROB can find the global optimal solution. The second algorithm is the gradient method in the price space (grad-PRICE). The third algorithm considers the pricing problem without network diffusion, that is, the price is determined according to the standard logit model. We still refer to it as the model misspecification (MM) scheme.

For the public price case, we also test three different algorithms. However, in this case, the pricing problem cannot be considered in the adoption probability space. Instead, we use a grid search (GS) to find an upper-bound solution for the problem. Specifically, we divide the price into grids of tolerance ξ. For each price p, we upper bound the profit with $(p-\xi)\cdot\sum_{i\in V}\mu_i(p)$. The other two algorithms, grad-PRICE and MM, as discussed above, are applied here.

For both scenarios, we test on a real-world network—*Amherst41* as introduced in Section 4.4.2. For each diffusion instance, we set the price sensitivity as $\alpha=0.1$, the network effect strength as $\beta=0.5$, and the intrinsic value $v_i \overset{i.i.d.}{\sim} \mathcal{U}(0,4)$. In Figures 8 and 9, we plot the realized profit of three algorithms and the relative profit loss against different values of the network effect parameter β. The relative profit loss is compared with the optimal (upper bound) results from grad-PROB and GS, respectively. Furthermore, as remarked before, algorithm grad-PRICE involves the derivation of the gradient as Eqn. (15), which requires calculating the inverse of a $|V|\times|V|$ matrix. We resort to the second-order approximate gradient, $\widetilde{G}_2(\boldsymbol{p})$, as given in Eqn. (16).

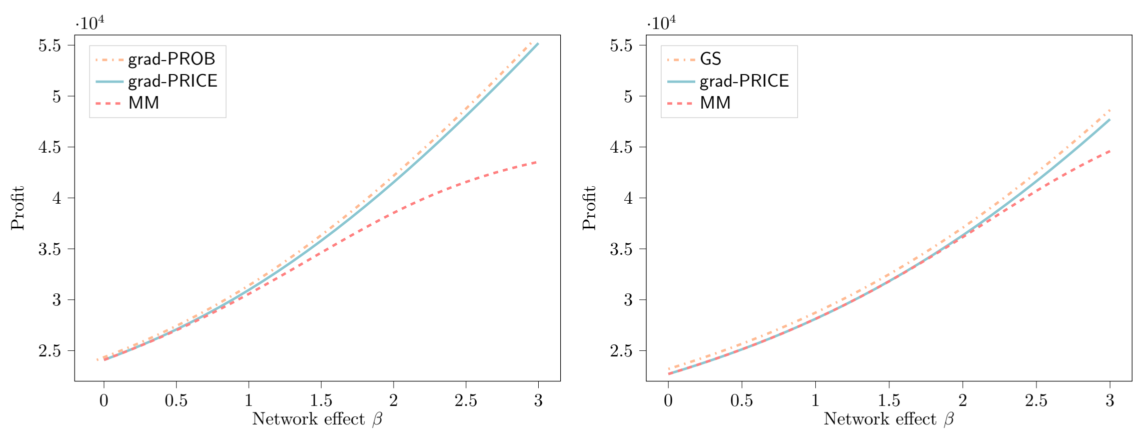


Figure 8 Realized profit versus network effect. Left: with price discrimination; Right: without price discrimination (The curve of grad-PROB coincides with grad-PRICE in the left figure. In order to make them identifiable in the figures, we shift the grad-PROB to the left by 0.05.)

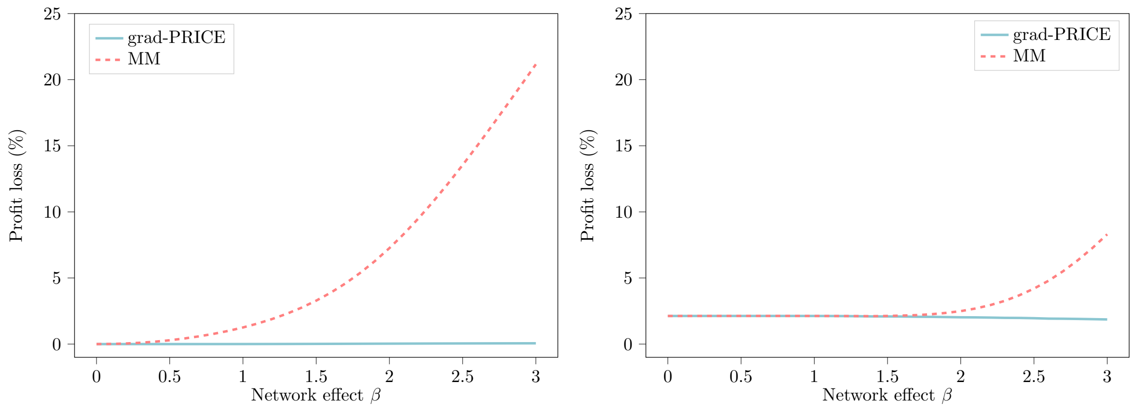


Figure 9 Profit loss compared with grad-PROB/GS versus network effect. Left: with price discrimination; Right: without price discrimination

We offer several observations from these two figures. First, grad-PRICE obtains a near-optimal solution in the case of price discrimination. This hints that we can use grad-PRICE to gain high-quality results in the general pricing setting when grad-PROB is not applicable. Second, there is a significant performance degradation of MM when the network effect is large. When $\beta = 3$, the relative profit loss reaches 21.16% and 8.30% if the network effect is ignored, respectively. Third, comparing these two scenarios, we find that pricing discrimination can significantly increase the total profit, especially when the network effect is large.

Furthermore, we compare the performance of these algorithms on more instances in terms of their execution time and the quality of solutions. We assume parameters $(N, \theta, v_{\max}) \in \{100, 1000\} \times \{0.1, 0.9\} \times \{4, 10\}$, where $v_{\max}$ is a parameter representing the range of intrinsic value v_i. Specifically, we assume v_i is i.i.d. sampled from $\mathcal{U}(0, v_{\max})$. The numerical results for the two scenarios are shown in Tables 2 and 3. The grad-PRICE approach derives high-quality solutions in both scenarios. We notice that in the perfect price discrimination case, the run time for grad-PROB is less than grad-PRICE, although the margin is not too large and the run times of the two algorithms are on a similar scale. The profit difference between these two approaches is quite small, uniformly smaller than 0.2%. For the public price case, we set the tolerance of the grid search to be 0.5 within the range $[0, 100]$. grad-PRICE runs much faster than the grid search with a performance loss of up to 2%. The performance of MM remains poor across the two scenarios in this experiment, suggesting that the loss from ignoring network effects can be detrimental. In summary, our main message through the numerical experiments is twofold. First, it is important to incorporate the network effect into operational problems. The gain from doing so can be significant. Second, we advocate grad-PRICE as a practical method for price optimization. With our approximate gradient expression tailored to the network setting as in Eqn. (16), grad-PRICE becomes a competitive price optimization technique. It can be efficiently implemented in various practical scenarios to find high-quality price solutions.

Table 2 Numerical results of pricing problem for randomly generated instances (perfect price discrimination)

Parameters $(N, \theta, v_{\max})$	grad-PROB	grad-PRICE				MM			
	time (s)	time (s)	profit loss (%)			time (s)	profit loss (%)		
			min	mean	max		min	mean	max
(100,0.1,4)	0.035	0.115	0.064	0.079	0.100	0.007	19.0340	20.126	20.971
(100,0.1,10)	0.095	0.095	0.005	0.009	0.013	0.006	18.475	19.208	19.766
(100,0.9,4)	0.025	0.131	0.066	0.089	0.120	0.006	18.300	19.724	20.874
(100,0.9,10)	0.083	0.093	0.006	0.009	0.013	0.006	18.241	19.067	19.200
(10,000,0.1,4)	11.041	81.223	0.009	0.009	0.009	7.753	19.008	19.082	19.138
(10,000,0.1,10)	11.062	81.191	0.009	0.009	0.009	7.727	19.019	19.074	19.140
(10,000,0.9,4)	76.184	151.138	0.085	0.087	0.090	8.363	19.654	19.790	19.901
(10,000,0.9,10)	78.501	160.918	0.009	0.009	0.009	7.967	19.011	19.070	19.148

Table 3 Numerical results of pricing problem for randomly generated instances (public price)

Parameters $(N, \theta, v_{\max})$	GS	grad-PRICE				MM			
	time (s)	time (s)	profit loss (%)			time (s)	profit loss (%)		
			min	mean	max		min	mean	max
(100,0.1,4)	0.652	0.059	1.728	1.871	2.068	0.005	4.102	7.837	10.115
(100,0.1,10)	0.687	0.093	0.832	0.964	1.134	0.007	0.891	1.333	2.505
(100,0.9,4)	0.685	0.062	1.698	1.857	2.071	0.005	5.762	8.137	11.547
(100,0.9,10)	0.720	0.096	0.833	0.957	1.159	0.007	0.924	1.395	3.040
(10,000,0.1,4)	137.674	27.147	1.827	1.860	1.866	0.014	7.523	7.880	8.259
(10,000,0.1,10)	151.969	60.072	0.952	0.968	0.984	0.047	0.973	1.024	1.120
(10,000,0.9,4)	1070.717	91.743	1.835	1.860	1.866	0.024	7.634	7.879	8.294
(10,000,0.9,10)	1203.320	170.557	0.951	0.966	0.984	0.042	0.984	1.028	1.106

6. Conclusion

In this study, we focus on nonprogressive diffusion in the social network. In the nonprogressive setting, people can withdraw their previous decision in accordance with a change in the social environment. We tide over the issues of the lack of a general modeling framework and efficient algorithms in the previous nonprogressive diffusion studies. Specifically, we build a general nonprogressive diffusion model that is agent-based, considers the local network effect, and can be adapted to many utility models. We propose a fixed-point proxy that can accurately and efficiently approximate the limiting adoption probability for all the agents under the diffusion model. Based on the fixed-point approximation, we investigate the conventional optimization problems. Finally, we use comprehensive numerical experiments to show the power of the fixed-point in different settings.

We also view one of our contributions as proposing a novel approach to study the long-run behavior of the agents in networks in stochastic settings. In particular, there are several directions for future research, in which our method seems readily extendable. First, the adoptions may not change in each period but last for several periods in practice (e.g., a user needs to subscribe to Netflix for at least one month). It would be interesting to investigate how we can represent the limiting behavior in this scenario. Second, this work only considers a binary-choice case where each agent only decides to adopt or not. It is worth investigating whether similar results can be extended to a multiple-choice case (e.g., not to subscribe, to subscribe to a normal membership, or to subscribe to a premium membership). Finally, the local network effect is captured by the average adoption of the in-neighbors in our model. It is promising to consider the weighted average of in-neighbor adoptions where the network effect is asymmetric.

References

Acemoğlu, D., Como, G., Fagnani, F., and Ozdaglar, A. (2013). Opinion fluctuations and disagreement in social networks. Mathematics of Operations Research, 38(1):1–27.

Afeche, P., Liu, Z., and Maglaras, C. (2018). Ride-hailing networks with strategic drivers: The impact of platform control capabilities on performance. Rotman School of Management Working Paper, (3120544):18–19.

Agrawal, S., Yin, S., and Zeevi, A. (2021). Dynamic pricing and learning under the bass model. In Proceedings of the 22nd ACM Conference on Economics and Computation, pages 2–3.

Anari, N., Ehsani, S., Ghodsi, M., Haghpanah, N., Immorlica, N., Mahini, H., and Mirrokni, V. S. (2010). Equilibrium pricing with positive externalities. In International Workshop on Internet and Network Economics, pages 424–431. Springer.

Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). Who's who in networks. wanted: The key player. Econometrica, 74(5):1403–1417.

Bapna, R. and Umyarov, A. (2015). Do your online friends make you pay? a randomized field experiment on peer influence in online social networks. Management Science, 61(8):1902–1920.

Baron, O., Hu, M., and Malekian, A. (2022). Revenue volatility under uncertain network effects. Operations Research.

Bass, F. M. (1969). A new product growth for model consumer durables. Management science, 15(5):215–227.

Benaïm, M. and Weibull, J. W. (2003). Deterministic approximation of stochastic evolution in games. Econometrica, 71(3):873–903.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. Journal of mathematical sociology, 2(1):113–120.

Candogan, O., Bimpikis, K., and Ozdaglar, A. (2012). Optimal pricing in networks with externalities. Operations Research, 60(4):883–905.

Chan, T.-H. H., Ning, L., and Zhang, Y. (2020). Influence maximization under the non-progressive linear threshold model. In International Workshop on Frontiers in Algorithmics, pages 37–48. Springer.

Chen, N. and Chen, Y.-J. (2021). Duopoly competition with network effects in discrete choice models. Operations Research, 69(2):545–559.

Chen, W., Wang, Y., and Yang, S. (2009). Efficient influence maximization in social networks. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 199–208.

Chen, W., Yuan, Y., and Zhang, L. (2010). Scalable influence maximization in social networks under the linear threshold model. In 2010 IEEE international conference on data mining, pages 88–97. IEEE.

Chen, Y. and Shi, C. (2022). Network revenue management with online inverse batch gradient descent method. Available at SSRN 3331939.

Cheng, S., Shen, H., Huang, J., Zhang, G., and Cheng, X. (2013). Staticgreedy: solving the scalability-accuracy dilemma in influence maximization. In Proceedings of the 22nd ACM international conference on Information & Knowledge Management, pages 509–518.

Datareportal (2022). Digital 2022: Global overview report. https://datareportal.com/reports/digital-2022-global-overview-report. Accessed: 2022-06-27.

Dewan, S., Ho, Y.-J., and Ramaprasad, J. (2017). Popularity or proximity: Characterizing the nature of social influence in an online music community. Information Systems Research, 28(1):117–136.

Drakopoulos, K. and Zheng, F. (2017). Network effects in contagion processes: Identification and control. Columbia Business School Research Paper, (18-8).

Du, C., Cooper, W. L., and Wang, Z. (2016). Optimal pricing for a multinomial logit choice model with network effects. Operations Research, 64(2):441–455.

Du, C., Cooper, W. L., and Wang, Z. (2018). Optimal worst-case pricing for a logit demand model with network effects. Operations Research Letters, 46(3):345–351.

Feng, Q., Li, C., Lu, M., and Shanthikumar, J. G. (2022). Implementing environmental and social responsibility programs in supply networks through multiunit bilateral negotiation. Management Science, 68(4):2579–2599.

Gallego, G. and Wang, R. (2014). Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities. Operations Research, 62(2):450–461.

Ghosh, R. and Lerman, K. (2012). Rethinking centrality: the role of dynamical processes in social network analysis. arXiv preprint arXiv:1209.4616.

Goldenberg, J., Libai, B., and Muller, E. (2001). Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing letters, 12(3):211–223.

Golrezaei, N., Jaillet, P., and Liang, J. C. N. (2020). No-regret learning in price competitions under consumer reference effects. Advances in Neural Information Processing Systems, 33:21416–21427.

Gopalakrishnan, M., Zhang, H., and Zhang, Z. (2022). Multiproduct pricing under the multinomial logit model with local network effects. Decision Sciences.

Goyal, A., Lu, W., and Lakshmanan, L. V. (2011). Celf++ optimizing the greedy algorithm for influence maximization in social networks. In Proceedings of the 20th international conference companion on World wide web, pages 47–48.

Granovetter, M. (1978). Threshold models of collective behavior. American journal of sociology, 83(6):1420–1443.

Hogben, L. (2006). Handbook of linear algebra. CRC press.

Horst, U. and Scheinkman, J. A. (2006). Equilibria in systems of social interactions. Journal of Economic Theory, 130(1):44–77.

Hu, M., Wang, Z., and Feng, Y. (2020). Information disclosure and pricing policies for sales of network goods. Operations Research, 68(4):1162–1177.

Jackson, M. O. (2010). Social and economic networks. Princeton university press.

Jackson, M. O., Lin, Z., and Yu, N. N. (2020). Adjusting for peer-influence in propensity scoring when estimating treatment effects. Available at SSRN 3522256.

Jackson, M. O. and Yariv, L. (2007). Diffusion of behavior and equilibrium properties in network games. American Economic Review, 97(2):92–98.

Kempe, D., Kleinberg, J., and Tardos, É. (2003). Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 137–146.

Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character, 115(772):700–721.

Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., and Glance, N. (2007). Cost-effective outbreak detection in networks. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 420–429.

Lewis, K. and Kaufman, J. (2018). The conversion of cultural tastes into social network ties. American journal of sociology, 123(6):1684–1742.

Li, H. (2020). Optimal pricing under diffusion-choice models. Operations Research, 68(1):115–133.

Li, H. and Huh, W. T. (2011). Pricing multiple products with the multinomial logit and nested logit models: Concavity and implications. Manufacturing & Service Operations Management, 13(4):549–563.

Li, Y., Fan, J., Wang, Y., and Tan, K.-L. (2018). Influence maximization on social graphs: A survey. IEEE Transactions on Knowledge and Data Engineering, 30(10):1852–1872.

Lin, Y., Wang, M., Shen, Z.-J. M., Zhang, H., and Zhang, R. P. (2021). Content promotion for online content platforms with network diffusion effect. Available at SSRN 3863104.

Lu, Y., Jerath, K., and Singh, P. V. (2013). The emergence of opinion leaders in a networked online community: A dyadic model with time dynamics and a heuristic for fast estimation. Management Science, 59(8):1783–1799.

Ma, L., Krishnan, R., and Montgomery, A. L. (2015). Latent homophily or social influence? an empirical analysis of purchase within a social network. Management Science, 61(2):454–473.

Nemhauser, G. L., Wolsey, L. A., and Fisher, M. L. (1978). An analysis of approximations for maximizing submodular set functions—i. Mathematical programming, 14(1):265–294.

Nosrat, F., Cooper, W. L., and Wang, Z. (2021). Pricing for a product with network effects and mixed logit demand. Naval Research Logistics (NRL), 68(2):159–182.

Rheinboldt, W. C. (1998). Methods for solving systems of nonlinear equations. SIAM.

Robinson, B. and Lakhani, C. (1975). Dynamic price models for new-product planning. Management science, 21(10):1113–1122.

Rossi, R. A. and Ahmed, N. K. (2015). The network data repository with interactive graph analytics and visualization. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

Sadler, E. (2020). Diffusion games. American Economic Review, 110(1):225–70.

Schelling, T. C. (1978). Micromotives and macrobehavior. Technical report.

Shakarian, P., Bhatnagar, A., Aleali, A., Shaabani, E., Guo, R., et al. (2015). Diffusion in social networks. Springer.

Shen, S., You, M., and Ma, Y. (2017). Single-commodity stochastic network design under demand and topological uncertainties with insufficient data. Naval Research Logistics (NRL), 64(2):154–173.

Shriver, S. K., Nair, H. S., and Hofstetter, R. (2013). Social ties and user-generated content: Evidence from an online social network. Management Science, 59(6):1425–1443.

Song, J.-S. and Zipkin, P. (2009). Inventories with multiple supply sources and networks of queues with overflow bypasses. Management Science, 55(3):362–372.

Sridhar, A. and Kar, S. (2021). Mean-field approximation for stochastic population processes in networks under imperfect information. arXiv preprint arXiv:2101.09644.

Tang, Y., Xiao, X., and Shi, Y. (2014). Influence maximization: Near-optimal time complexity meets practical efficiency. In Proceedings of the 2014 ACM SIGMOD international conference on Management of data, pages 75–86.

Van Mieghem, P., Omic, J., and Kooij, R. (2008). Virus spread in networks. IEEE/ACM Transactions On Networking, 17(1):1–14.

Wang, C., Chen, W., and Wang, Y. (2012). Scalable influence maximization for independent cascade model in large-scale social networks. Data Mining and Knowledge Discovery, 25(3):545–576.

Wang, R. and Wang, Z. (2017). Consumer choice models with endogenous network effects. Management Science, 63(11):3944–3960.

Wasserman, S. (1994). Advances in social network analysis: Research in the social and behavioral sciences. Sage.

Xie, J. and Sirbu, M. (1995). Price competition and compatibility in the presence of positive demand externalities. Management science, 41(5):909–926.

Xie, T. and Wang, Z. (2020). Personalized assortment optimization under consumer choice models with local network effects. Available at SSRN 3788880.

Xu, H. (2018). Social interactions in large networks: A game theoretic approach. International Economic Review, 59(1):257–284.

Yang, N. and Zhang, R. P. (2022). Dynamic pricing and inventory management in the presence of online reviews. Production and Operations Management, 31(8):3180–3197.

Zeng, Z., Dai, H., Zhang, D., Zhang, H., Zhang, R. P., Xu, Z., and Shen, Z.-J. M. (2021). The impact of social nudges on user-generated content for social network platforms. Available at SSRN 3611571.

Zhou, C., Zhang, P., Zang, W., and Guo, L. (2015). On the upper bounds of spread for greedy algorithms in social network influence maximization. IEEE Transactions on Knowledge and Data Engineering, 27(10):2770–2783.

Online Appendices

Appendix A: Proofs in Section 3

The following lemma is useful in the proof of Proposition 1.

LEMMA 3. *For each agent $i \in V$, at least one of following value*

$$(a) \prod_{\boldsymbol{y}\in\{0,1\}^{|V|}} \mathbb{P}(Y_i(t)=1|\boldsymbol{Y}(t-1)=\boldsymbol{y}) \ or \ (b) \prod_{\boldsymbol{y}\in\{0,1\}^{|V|}} \mathbb{P}(Y_i(t)=0|\boldsymbol{Y}(t-1)=\boldsymbol{y})$$

are positive.

Proof of Lemma 3: When random noise $\epsilon_i(t)$ is not bounded on either side, it is obvious that the statement holds. In the following, we only consider the situation when $\epsilon_i(t))$ is with support on some bounded interval $[\underline{\epsilon}, \overline{\epsilon}]$.

If (a) is not positive, there exists $\boldsymbol{y} \in \{0,1\}^{|V|}$ such that $F_\epsilon(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} y_j}{n_i}) = 1$. Hence, we have

$$-v_i - \inf_{\boldsymbol{y}\in\{0,1\}^{|V|}} \beta\frac{\sum_{j\in\mathcal{N}_i} y_j}{n_i} = -v_i \geq \overline{\epsilon}$$

Consequently, we can derive the following inequality

$$-v_i - \beta > -v_i - \frac{1}{L} \geq -v_i - (\overline{\epsilon} - \underline{\epsilon}) \geq \underline{\epsilon}$$

where the first inequality follows from Assumption 2, the second inequality follows from property (ii) in Assumption 1. As a direct result, for any $\boldsymbol{y} \in \{0,1\}^{|V|}$, $F_\epsilon(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} y_j}{n_i}) > 0$ and the value of (b) is positive.

In conclusion, for each agent $i \in V$, at least one of (a) and (b) have positive value. □

Proof of Proposition 1: Our goal is to show that although the MC may not be irreducible, there is only one recurrent communication class and it is aperiodic.

We first show that there is only one recurrent communication class. Follow after Lemma 3, we construct $\boldsymbol{y}'$ as follows

$$y'_i = \begin{cases} 0 & \text{when } \prod_{\boldsymbol{y}\in\{0,1\}^{|V|}} \mathbb{P}(Y_i(t)=1|\boldsymbol{Y}(t-1)=\boldsymbol{y}) = 0 \\ 1 & \text{o.w.} \end{cases}$$

Hence, we have $\mathbb{P}(Y_i(t) = y'_i|\boldsymbol{Y}(t-1) = \boldsymbol{y}) > 0$ for all $i \in V$ and $\boldsymbol{y} \in \{0,1\}^{|V|}$. When given the previous adoption state, each agent $i \in V$ makes their decision independently. Consequently, $p(\boldsymbol{y},\boldsymbol{y}') = \prod_{i\in V}\mathbb{P}(Y_i(t) = y'_i|\boldsymbol{Y}(t-1) = \boldsymbol{y}) > 0$ holds for all $\boldsymbol{y} \in \{0,1\}^{|V|}$. It further implies that all states of this MC communicates with state $\boldsymbol{y}'$. As a result, the states that $\boldsymbol{y}'$ communicates with forms a recurrent communication class while other states are in the transient classes.

Further, we can notice that since $p(\boldsymbol{y}',\boldsymbol{y}') > 0$ also holds, state $\boldsymbol{y}'$ has period 1 which implies that the recurrent communication class is aperiodic.

In conclusion, this MC has a limiting distribution $\boldsymbol{\pi}$ that satisfies $\boldsymbol{\pi} = \boldsymbol{\pi}P$ and the limiting adoption probability of each agent is a linear transformation of $\boldsymbol{\pi}$ that follows Eqn. (3). □

Appendix B: Example and Proofs in Section 4

B.1. Diffusion instance example

In this example, we assume that the network effect parameter is $\beta = 3.5$ and the random noise $\epsilon_i(t) \overset{i.i.d.}{\sim}$ Logistic$(0,1)$. We construct a diffusion instance with 10 nodes. Table 4 shows the information of this instance as well as the numerical results of adoption probability.

Table 4 Diffusion instance information and adoption probability

Node	Degree	Intrinsic value v	p^*	μ^*	FPA error	p^{MM}	MM error
0	5	-1.7064	0.5139	0.5292	0.0154	0.1536	-0.3602
1	7	-1.2453	0.5918	0.6069	0.0150	0.2235	-0.3683
2	4	-0.8789	0.6314	0.6524	0.0210	0.2934	-0.3380
3	4	-3.9454	0.1435	0.1221	-0.0215	0.0190	-0.1246
4	3	-0.0822	0.7811	0.8219	0.0408	0.4795	-0.3017
5	5	-3.4441	0.1943	0.1731	-0.0212	0.0309	-0.1633
6	3	-0.2877	0.7350	0.7755	0.0405	0.4286	-0.3065
7	2	-2.9084	0.3284	0.2849	-0.0434	0.0517	-0.2766
8	2	-1.2859	0.6709	0.7646	0.0936	0.2166	-0.4544
9	1	-0.6963	0.7440	0.8786	0.1346	0.3326	-0.4114

B.2. Proofs in Section 4

Proof of Theorem 2: We will first show property (i) and then proof property (ii) and (iii) by showing that $\boldsymbol{h}(\cdot)$ is a contraction mapping.

Proof of (i): When $\boldsymbol{a} \leq \boldsymbol{b}$, we have $\sum_{j\in\mathcal{N}_i} a_j \leq \sum_{j\in\mathcal{N}_i} b_j$ for all $i \in V$. Since CDF $F_\epsilon(\cdot)$ is monotonically increasing, if $\boldsymbol{a} \leq \boldsymbol{b}$,

$$1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} a_j}{n_i}\right) \leq 1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} b_j}{n_i}\right)$$

for all $i \in V$, which implies $h(\boldsymbol{a}) \leq h(\boldsymbol{b})$.

Proof of (ii) and (iii):

It is trivial that $h(\cdot)$ maps $\mathbb{R}^{|V|}$ to itself.

Consider the Jacobian matrix of $\boldsymbol{h}(\boldsymbol{\mu})$, for all $\boldsymbol{\mu} \in \mathbb{R}^{|V|}$,

$$\frac{\partial h(\boldsymbol{\mu})_i}{\partial \mu_j} = \begin{cases} 0, & j \notin \mathcal{N}_i \\ \dfrac{\beta}{n_i}\dfrac{\partial F_\epsilon\left(-v_i - \beta\frac{\sum_{j'\in\mathcal{N}_i}\mu_{j'}}{n_i}\right)}{\partial\left(-v_i - \beta\frac{\sum_{j'\in\mathcal{N}_i}\mu_{j'}}{n_i}\right)}, & j \in \mathcal{N}_i. \end{cases}$$

By property (ii) of Assumption 1, we can have $\left|\frac{\partial h(\boldsymbol{\mu})_i}{\partial \mu_j}\right| \leq \frac{\beta L}{n_i}$ for all $j \in \mathcal{N}_i$. Therefore, the ∞-norm of $d\boldsymbol{h}(\boldsymbol{\mu})/d\boldsymbol{\mu}$ can be upper bounded as

$$\left\|\frac{d\boldsymbol{h}(\boldsymbol{\mu})}{d\boldsymbol{\mu}}\right\|_\infty = \max_{i\in V}\sum_{j\in V}\left|\frac{\partial h(\boldsymbol{\mu})_i}{\partial \mu_j}\right| \leq \max_{i\in V} n_i \frac{\beta L}{n_i} = \beta L < 1$$

where the last inequality follows from Assumption 2.

Thus, for all $\boldsymbol{\mu} \in \mathbb{R}^N$, we have $\|d\boldsymbol{h}(\boldsymbol{\mu})/d\boldsymbol{\mu}\|_\infty < 1$.

It then implies that $h(\boldsymbol{\mu})$ is a contraction mapping. By contraction mapping theorem, we conclude the proof. □

LEMMA 4 **(Variance of function).** *Let X be a random variable. If $g(\cdot)$ is a L-Lipschitz continuous function, the following inequality holds.*

$$Var(g(X)) \le L^2\, Var(X).$$

Proof of Lemma 4:

$$\mathrm{Var}(g(X)) = \mathrm{Var}(g(X) - g(\mathbb{E}[X])) \le \mathbb{E}[(g(X) - g(\mathbb{E}[X]))^2] \le \mathbb{E}[L^2(X - \mathbb{E}[X])^2] = L^2 \mathrm{Var}(X)$$

where the first inequality follows since for any random variable Y, $\mathrm{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$, and the last inequality follows from the L-Lipschitz continuity. □

Proof of Lemma 1: Using law of total covariance, we can decompose $\mathrm{Cov}(Y_i(t), Y_{i'}(t))$ into two parts as

$$\mathrm{Cov}(Y_i(t), Y_{i'}(t)) = \mathbb{E}_{\boldsymbol{Y}(t-1)}\left[\mathrm{Cov}(Y_i(t), Y_{i'}(t)|\boldsymbol{Y}(t-1))\right] + \mathrm{Cov}_{\boldsymbol{Y}(t-1)}\left(\mathbb{E}[Y_i(t)|\boldsymbol{Y}(t-1)], \mathbb{E}[Y_{i'}(t)|\boldsymbol{Y}(t-1)]\right).$$

The first term $\mathbb{E}_{\boldsymbol{Y}(t-1)}\left[\mathrm{Cov}(Y_i(t), Y_{i'}(t)|\boldsymbol{Y}(t-1))\right]$ is always 0. By applying law of total covariance again, we have

$$\mathrm{Cov}(Y_i(t), Y_{i'}(t)|\boldsymbol{Y}(t-1)) = \mathbb{E}_{\boldsymbol{\epsilon}_t}[\mathrm{Cov}(Y_i(t), Y_{i'}(t))|\boldsymbol{Y}(t-1), \boldsymbol{\epsilon}_t] + \mathrm{Cov}(\mathbb{E}_{\boldsymbol{\epsilon}(t)}[Y_i(t)|\boldsymbol{\epsilon}(t)], \mathbb{E}_{\boldsymbol{\epsilon}(t)}[Y_{i'}(t)|\boldsymbol{\epsilon}(t)]|\boldsymbol{Y}(t-1)).$$

The former term vanishes because $Y_i(t)$ and $Y_{i'}(t)$ are deterministic when given $\boldsymbol{Y}(t-1), \boldsymbol{\epsilon}_t$. The latter term is also zero since $\epsilon_i(t)$ and $\epsilon_{i'}(t)$ are independent of each other.

We then show that the second term $\mathrm{Cov}_{\boldsymbol{Y}(t-1)}\left(\mathbb{E}[Y_i(t)|\boldsymbol{Y}(t-1)], \mathbb{E}[Y_{i'}(t)|\boldsymbol{Y}(t-1)]\right)$ can be bounded iteratively as

$$\begin{aligned}
&\mathrm{Cov}_{\boldsymbol{Y}(t-1)}\left(\mathbb{E}[Y_i(t)|\boldsymbol{Y}(t-1)], \mathbb{E}[Y_{i'}(t)|\boldsymbol{Y}(t-1)]\right) \\
=&\mathrm{Cov}\left(1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{n_i}\right), 1 - F_\epsilon\left(-v_{i'} - \beta\frac{\sum_{j'\in\mathcal{N}_{i'}} Y_{j'}(t-1)}{n_i'}\right)\right) \\
\le&\sqrt{\mathrm{Var}\left(F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{n_i}\right)\right)\mathrm{Var}\left(F_\epsilon\left(-v_{i'} - \beta\frac{\sum_{j'\in\mathcal{N}_{i'}} Y_{j'}(t-1)}{n_{i'}}\right)\right)} && \text{(17a)} \\
\le&\sqrt{\frac{(L\beta)^2}{n_i^2}\mathrm{Var}\left(\sum_{j\in\mathcal{N}_i} Y_j(t-1)\right)\frac{(L\beta)^2}{n_{i'}^2}\mathrm{Var}\left(\sum_{j'\in\mathcal{N}_{i'}} Y_{j'}(t-1)\right)} && \text{(17b)} \\
=&\frac{(L\beta)^2}{n_i n_{i'}}\sqrt{\mathrm{Var}\left(\sum_{j\in\mathcal{N}_i} Y_j(t-1)\right)\mathrm{Var}\left(\sum_{j'\in\mathcal{N}_{i'}} Y_{j'}(t-1)\right)}
\end{aligned}$$

where (17a) follows from Cauchy–Schwarz inequality and (17b) follows from Lemma 4. We show by induction that for every time $t \ge 1$,

$$\mathrm{Cov}(Y_i(t), Y_{i'}(t)) \le \frac{0.25\sum_{\tau=1}^{t-1}(L\beta)^{2\tau}}{N_{\min}}.$$

$\boldsymbol{t=1}$: We give some basic facts at the first iteration which will support the subsequent proof. Since users act independently when all the users are unadopted at the very beginning, it is obvious that $\mathrm{Cov}(Y_{i,1}, Y_{i',1}) = 0$

for all $i, i' \in V$. Incorporating the trivial fact that the variance of a binary random variable can be no larger than 0.25, that is, $\mathrm{Var}(Y_i(t)) \leq 0.25$ for all $i \in V$ and $t \geq 0$, we get

$$\mathrm{Var}\left(\sum_{j\in\mathcal{N}_i} Y_j(1)\right) = \sum_{j\in\mathcal{N}_i} \mathrm{Var}\left(Y_j(1)\right) + \sum_{j,j'\in\mathcal{N}_i, j\neq j'} \mathrm{Cov}\left(Y_j(1), Y_{j'}(1)\right) \leq 0.25 n_i. \tag{18}$$

***t* = 2**: This is the base case for the statement. Following inequality (17) and (18), we have

$$\mathrm{Cov}(Y_i(2), Y_{i'}(2)) \leq \frac{(L\beta)^2}{n_i n_{i'}} \sqrt{\mathrm{Var}\left(\sum_{j\in\mathcal{N}_i} Y_j(1)\right) \mathrm{Var}\left(\sum_{j'\in\mathcal{N}_{i'}} Y_{j'}(1)\right)} \leq \frac{0.25(L\beta)^2}{\sqrt{n_i n_{i'}}} \leq \frac{0.25(L\beta)^2}{N_{\min}}.$$

Assume *t* = *s*: The induction hypothesis holds such that

$$\mathrm{Cov}(Y_{i,s}, Y_{i',s}) \leq \frac{0.25 \sum_{\tau=1}^{s-1} (L\beta)^{2\tau}}{N_{\min}}.$$

We can derive that for all $i \in V$,

$$\mathrm{Var}\left(\sum_{j\in\mathcal{N}_i} Y_j(s)\right) = \sum_{j\in\mathcal{N}_i} \mathrm{Var}\left(Y_j(s)\right) + \sum_{j,j'\in\mathcal{N}_i, j\neq j'} \mathrm{Cov}\left(Y_j(s), Y_{j'}(s)\right) \leq 0.25 n_i + n_i(n_i - 1)\frac{0.25 \sum_{\tau=1}^{s-1} (L\beta)^{2\tau}}{N_{\min}}. \tag{19}$$

Show *t* = *s* + 1: Following inequality (19), we have

$$\begin{aligned}
\mathrm{Cov}(y_i(s+1), y_{i'}(s+1)) &\leq \frac{(L\beta)^2}{n_i n_{i'}} \sqrt{\mathrm{Var}\left(\sum_{j\in\mathcal{N}_i} y_j(s)\right) \mathrm{Var}\left(\sum_{j'\in\mathcal{N}_j} y_{j'}(s)\right)} \\
&\leq (L\beta)^2 \sqrt{\left(\frac{0.25}{n_i} + \frac{0.25 \sum_{\tau=1}^{s-1} (L\beta)^{2\tau}}{N_{\min}}\right)\left(\frac{0.25}{n_{i'}} + \frac{0.25 \sum_{\tau=1}^{s-1} (L\beta)^{2\tau}}{N_{\min}}\right)} \\
&\leq 0.25 (L\beta)^2 \left(\frac{1 + \sum_{\tau=1}^{s-1} (L\beta)^{2\tau}}{N_{\min}}\right) \\
&= \frac{0.25 \sum_{\tau=1}^{s} (L\beta)^{2\tau}}{N_{\min}}.
\end{aligned}$$

Thus, the statement also holds true when $t = s+1$, establishing the inductive step.

Finally, we have

$$\mathrm{Cov}(Y_i(t), Y_{i'}(t)) \leq \frac{0.25 \sum_{\tau=1}^{t-1} (L\beta)^{2\tau}}{N_{\min}} \leq \lim_{t\to+\infty} \frac{0.25 \sum_{\tau=1}^{t-1} (L\beta)^{2\tau}}{N_{\min}} = \lim_{t\to+\infty} \frac{0.25(L\beta)^2[1-(L\beta)^{2t}]}{(1-(L\beta)^2)N_{\min}} = \frac{0.25(L\beta)^2}{1-(L\beta)^2}\frac{1}{N_{\min}}$$

where we assume $0 < L\beta < 1$ according to Asssumption 2. □

Proof of Lemma 2: Let $R_i(t-1) = \beta \frac{\sum_{j\in\mathcal{N}_i} q_j(t-1) - Y_j(t-1)}{n_i}$ for any $i \in V$ and $t \geq 1$, the adoption probability of agent i at t can be written as

$$\begin{aligned}
q_i(t) = \mathbb{E}_{\boldsymbol{Y}(t-1)}\left[\mathbb{E}[y_i(t) | \boldsymbol{Y}(t-1)]\right] &= \mathbb{E}_{\boldsymbol{Y}(t-1)}\left[1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{n_i}\right)\right] \\
&= 1 - \mathbb{E}_{\boldsymbol{Y}(t-1)}\left[F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{n_i} + R_i(t-1)\right)\right].
\end{aligned}$$

Since $F_\epsilon(\cdot)$ is a L-Lipschitz continuous function, we have

$$\left| F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{n_i} + R_i(t-1)\right) - F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{n_i}\right) \right| \leq L |R_i(t-1)|.$$

To bound the expectation of $L|R_i(t-1)|$, we have

$$\mathbb{E}[L|R_i(t-1)|] = \sqrt{(\mathbb{E}[L|R_i(t-1)|])^2} \leq \sqrt{\mathbb{E}[(LR_i(t-1))^2]} = \sqrt{\mathrm{Var}(LR_i(t-1)) + (\mathbb{E}[LR_i(t-1)])^2}$$

where the inequality follows from Jensen's inequality. By applying Proposition 1, we can further derive

$$\begin{aligned}
\mathrm{Var}(LR_i(t-1)) &= \frac{(L\beta)^2}{n_i^2}\mathrm{Var}\left(\sum_{j\in\mathcal{N}_i} Y_j(t-1)\right) \\
&= \frac{(L\beta)^2}{n_i^2}\left[\sum_{j\in\mathcal{N}_i}\mathrm{Var}(Y_j(t-1)) + \sum_{j,j'\in\mathcal{N}_i, j\neq j'}\mathrm{Cov}(Y_j(t-1), Y_{j'}(t-1))\right] \\
&\leq \frac{(L\beta)^2}{n_i^2}\left[0.25n_i + n_i(n_i-1)\frac{0.25(L\beta)^2}{1-(L\beta)^2}\frac{1}{N_{\min}}\right] \\
&\leq \frac{0.25(L\beta)^2}{1-(L\beta)^2}\frac{1}{N_{\min}}
\end{aligned}$$

and

$$\mathbb{E}[LR_i(t-1)] = \frac{L\beta}{n_i}\mathbb{E}\left[\sum_{j\in\mathcal{N}_i}(q_j(t-1) - Y_j(t-1))\right] = \frac{L\beta}{n_i}\sum_{j\in\mathcal{N}_i}(q_j(t-1) - \mathbb{E}[Y_j(t-1)]) = 0$$

which implies

$$\mathbb{E}[L|R_i(t-1)|] \leq \sqrt{\frac{0.25(L\beta)^2}{1-(L\beta)^2}\frac{1}{N_{\min}}} = \delta.$$

Therefore, we can have the following results,

$$\mathbb{E}\left[\left|F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{n_i} + R_i(t-1)\right) - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{n_i}\right)\right|\right] \leq \mathbb{E}\left[L|R_i(t-1)|\right] \leq \delta$$

By Jensen's inequality again, we have

$$\begin{aligned}
&\left|\mathbb{E}\left[F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{n_i} + R_i(t-1)\right) - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{n_i}\right)\right]\right| \\
\leq &\mathbb{E}\left[\left|F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{n_i} + R_i(t-1)\right) - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{n_i}\right)\right|\right]
\end{aligned}$$

which further leads to

$$1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{n_i}\right) - \delta \leq q_i(t) \leq 1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{n_i}\right) + \delta.$$

In conclusion, $\boldsymbol{h}_{-\delta}(\boldsymbol{q}(t-1)) \leq \boldsymbol{q}(t) \leq \boldsymbol{h}_\delta(\boldsymbol{q}(t-1))$. $\square$

Proof of Theorem 1: We first show by induction that for any diffusion instance, $\underline{\boldsymbol{\mu}}(t) \leq \boldsymbol{q}(t) \leq \overline{\boldsymbol{\mu}}(t)$ for each $t \geq 0$ on the trajectory.

$\boldsymbol{t=0}$: By definition, we have $\underline{\boldsymbol{\mu}}(0) = \boldsymbol{q}(0) = \overline{\boldsymbol{\mu}}(0)$.

Assume $\boldsymbol{t=s}$: The induction hypothesis holds such that $\underline{\boldsymbol{\mu}}(0) \leq \boldsymbol{q}(0) \leq \overline{\boldsymbol{\mu}}(0)$.

Show $\boldsymbol{t=s+1}$: We have

$$\underline{\boldsymbol{\mu}}_{s+1} = \boldsymbol{h}_{-\delta}(\underline{\boldsymbol{\mu}}(s)) \leq \boldsymbol{h}_{-\delta}(\boldsymbol{q}(s)) \leq \boldsymbol{q}(s+1) \leq \boldsymbol{h}_\delta(\boldsymbol{q}(s)) \leq \boldsymbol{h}_\delta(\overline{\boldsymbol{\mu}}(s)) = \overline{\boldsymbol{\mu}}(s+1)$$

where the first and last inequalities follow from property (i) in Theorem 2 while the other two inequalities follow from Lemma 2.

By contraction mapping theorem, we know that $\underline{\boldsymbol{\mu}}(t)$ (resp. $\overline{\boldsymbol{\mu}}(t)$) converges to $\underline{\boldsymbol{\mu}}^*$ (resp. $\overline{\boldsymbol{\mu}}^*$) where $\underline{\boldsymbol{\mu}}^*$ (resp. $\overline{\boldsymbol{\mu}}^*$) is the fixed-point solution for $\boldsymbol{h}_{-\delta}(\underline{\boldsymbol{\mu}}^*) = \underline{\boldsymbol{\mu}}^*$ (resp. $\boldsymbol{h}_{\delta}(\overline{\boldsymbol{\mu}}^*) = \overline{\boldsymbol{\mu}}^*$). Thus, the following result holds,

$$\underline{\boldsymbol{\mu}}^* \leq \boldsymbol{q}^* \leq \overline{\boldsymbol{\mu}}^* \tag{20}$$

Similarly, we can also have

$$\underline{\boldsymbol{\mu}}^* \leq \boldsymbol{\mu}^* \leq \overline{\boldsymbol{\mu}}^* \tag{21}$$

We then show the bound of the infinity norm of the difference between $\overline{\boldsymbol{\mu}}^*$ and $\underline{\boldsymbol{\mu}}^*$. By definition, we have

$$\overline{\boldsymbol{\mu}}^* - \underline{\boldsymbol{\mu}}^* = \boldsymbol{h}(\overline{\boldsymbol{\mu}}^*) - \boldsymbol{h}(\underline{\boldsymbol{\mu}}^*) + 2\delta \boldsymbol{e}$$

Let $\Delta\boldsymbol{\mu} = \overline{\boldsymbol{\mu}}^* - \underline{\boldsymbol{\mu}}^*$, for all $i \in V$,

$$\Delta\mu_i \leq L\beta \left| \frac{\sum_{j\in\mathcal{N}_i} \overline{\mu}_j^* - \underline{\mu}_j^*}{n_i} \right| + 2\delta = \frac{L\beta}{n_i} \sum_{j\in\mathcal{N}_i} \left| \overline{\mu}_j^* - \underline{\mu}_j^* \right| + 2\delta \leq L\beta \|\Delta\boldsymbol{\mu}\|_\infty + 2\delta$$

where the first inequality comes from L-Lipschitz continuity and the last inequality follows from the definition of infinity norm. Thus, we have

$$\|\Delta\boldsymbol{\mu}\|_\infty = \max_{i\in V} |\Delta\boldsymbol{\mu}_i| \leq L\beta \|\Delta\boldsymbol{\mu}\|_\infty + 2\delta$$

which implies $\|\Delta\boldsymbol{\mu}\|_\infty \leq \frac{2}{1-L\beta}\sqrt{\frac{(L\beta)^2}{1-(L\beta)^2}\frac{1}{N_{\min}}}$.

Combining (20) and (21), for an arbitrary diffusion instance with network neighbor size $N_{\min}$, we can have the following results.

$$\|\boldsymbol{q}^* - \boldsymbol{\mu}^*\|_\infty \leq \|\Delta\boldsymbol{\mu}\|_\infty \leq \frac{2}{1-L\beta}\sqrt{\frac{(L\beta)^2}{1-(L\beta)^2}\frac{1}{N_{\min}}}.$$

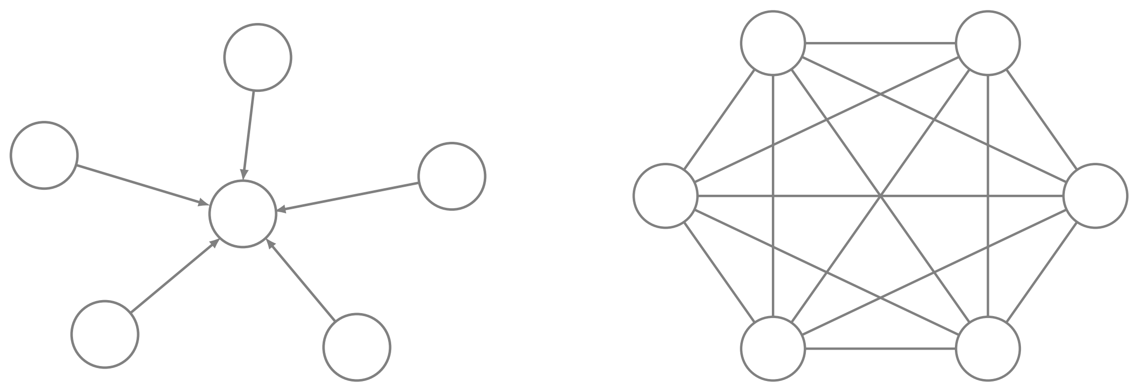
Therefore, we can conclude that $\|\boldsymbol{q}^* - \boldsymbol{\mu}^*\|_\infty \leq \frac{2}{1-L\beta}\sqrt{\frac{(L\beta)^2}{1-(L\beta)^2}\frac{1}{N_{\min}}}$. □

B.3. Numerical Experiments of FPA solution on Highly-structured Symmetric Networks

We choose two kinds of highly-structured symmetric networks, namely directed star network and complete network, for illustration. These networks have a simple and symmetric structure. We further assume all agents in the network have the same intrinsic values so that we are able to directly compute the limiting adoption probability or construct a simple Markov chain with fewer states.

Network instances:

- *Directed star network*. Star network consists of a central node and several surrounding nodes. Here, we consider a directed version where the edges only point from surrounding nodes to the central node. Figure 10a shows an example with network size $N = 6$.
- *Complete network*. Complete network is the network where all nodes are connected with each other. Figure 10b shows an example with network size $N = 6$.



(a) Directed star network (b) Complete network

Figure 10 Illustration of network structure. Left: directed star network, $N = 6$; Right: complete network, $N = 6$.

For directed star network, the actions of surrounding nodes are independent with each other. Therefore, we can directly derive the limiting adoption of central node as

$$p = \sum_{i=0}^{N-1} \binom{N-1}{i} (1 - F_\epsilon(-v))^i F_\epsilon(-v)^{N-1-i} \cdot \left[1 - F_\epsilon\left(-v - \beta \frac{i}{N-1}\right)\right]$$

For complete network, we use the number of adopted nodes as the state. The transition probability can be defined as

$$\begin{aligned} P(i,j) = \sum_{k=0}^{\min\{i,j\}} & \binom{i}{k} \left[1 - F_\epsilon\left(-v - \beta\frac{i-1}{N-1}\right)\right]^k F_\epsilon\left(-v - \beta\frac{i-1}{N-1}\right)^{i-k} \binom{n-i}{j-k} \\ & \cdot \left[1 - F_\epsilon\left(-v - \beta\frac{i}{N-1}\right)\right]^{j-k} F_\epsilon\left(-v - \beta\frac{i}{N-1}\right)^{n-i-j+k} \end{aligned}$$

The limiting adoption probability can hence be calculated after characterizing the limiting distribution of this MC.

The performance of the FPA scheme for each agent is measured by the percentage error (PE), given in the following equation.

$$\mathrm{PE} = \frac{\mu_i^* - q_i}{q_i} \cdot 100\%$$

For directed star networks, we only focus on the central node as the adoption of surrounding nodes are independent of all other agents. For complete networks, we can use the PE of an arbitrary node as it is the same amongst all nodes.

For both network structures, we explore the performance of the FPA scheme in two scenarios: a sequence of diffusion instances with different intrinsic values and another sequence with different network sizes. We set the network effect parameter to be $\beta = 1$ and generate the random noise $\epsilon_i(t) \overset{i.i.d.}{\sim} \text{Logistic}(0,1)$.

The accuracy with regard to intrinsic values. We choose the homogeneous intrinsic value v from -5 to 5 with an interval of 0.1. We test on the instances with network size $N \in \{10, 20, 30\}$. Figure 11 shows the PE of two network structures with different intrinsic values. Overall, all the instances have small absolute percentage error (less than 0.5%), illustrating the high accuracy of the FPA solution. We notice that the PE curves of different network structures possess similar shape while the exact values are slightly different.

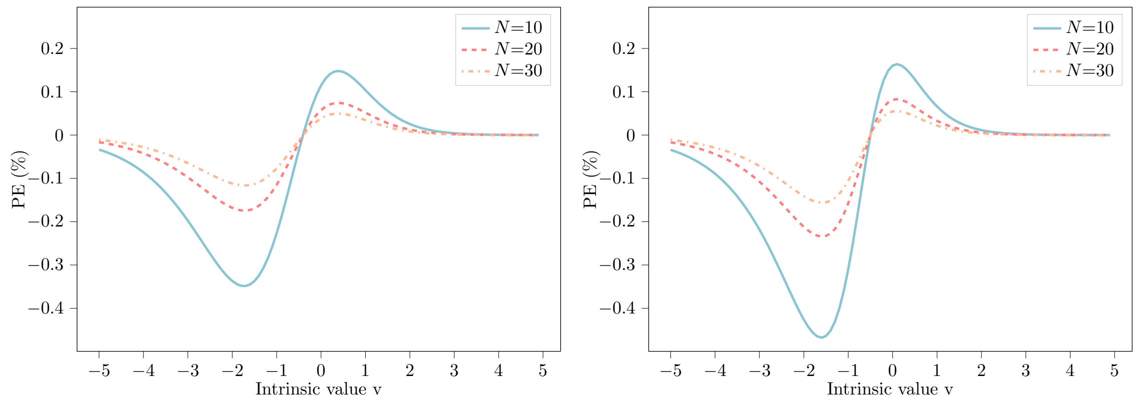


Figure 11 PE versus intrinsic value. Left: directed star network; Right: complete network.

In general, when the intrinsic values are small (resp. large) for all agents, the proxy tends to underestimate (resp. overestimate) the adoption probability. There exist two critical points at around $v=-1.7$ and $v=0.4$ where the PE arrives at an extreme. These two points exhibit the worst cases of the FPA scheme and the corresponding intrinsic values coincide with the region where CDF F_ϵ has the largest curvature.

The accuracy with regard to network size. We test on the instances with intrinsic values at the two critical points $v\in\{-1.7, 0.4\}$ mentioned before. We choose the network size N from 2 to 50. Figure 12 shows the PE of different network sizes. Regardless of the network structure and the intrinsic values, PE converges to

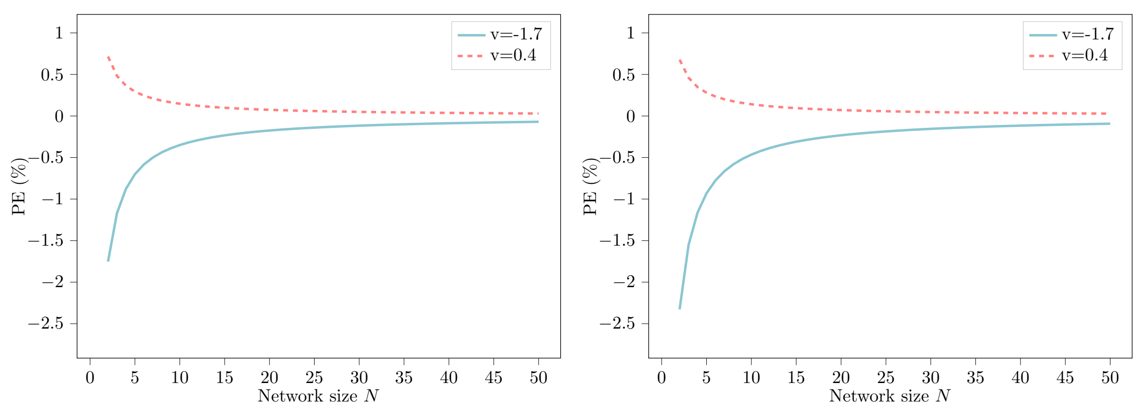


Figure 12 PE versus intrinsic value. Left: directed star network; Right: complete network.

0 rapidly when the network size (and minimum indegree) increases. This can be established theoretically, and we have formally showed this in Theorem 1. The results show that for the highly structured network structure, the FPA scheme shows excellent approximation quality and exhibits the asymptotic convergence when the network size increases.

B.4. Warm-up period

We run simulations to empirically show when the MC enters steady state so that the data samples can be gathered to calculate limiting adoption probability. Instead of focusing on the probability of each state of the MC, we use the average cumulative adoption proportion among the population as an indicator, given in the following equation.

$$\frac{1}{t}\sum_{\tau=1}^{T}\frac{1}{N}\sum_{i\in V}y_{i,\tau}$$

In Figure 13, we show that how the average cumulative adoption proportion changes with time. We test on 4 different diffusion instances. The network of each instance is a randomly sampled random graph $G(N,\theta)$ where we choose the network size from $N\in\{10,100,1000,10000\}$ and keep the probability of edge existence to be $\theta=0.1$. We set the network effect coefficient β to be 1, intrinsic value $v_i\overset{i.i.d.}{\sim}\mathcal{U}[-4,0]$ and random noise $\epsilon_i(t)\overset{i.i.d.}{\sim}\text{Logistic}(0,1)$.

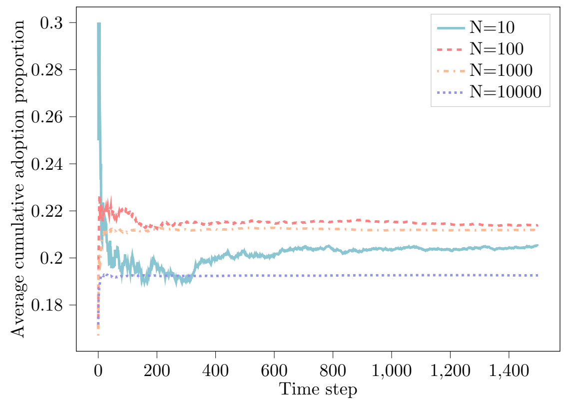


Figure 13 **Average cumulative adoption proportion versus MC time steps**

We can observe that after 1,000 time steps, all trajectories have entered the steady state. When N is large, the MC enters the steady state in fewer steps. We also test on other diffusion instances with different parameters, similar evidences can be observed. Therefore, we can conclude that it is sufficient to take the first 1,000 time steps as warm-up time period in our problem setting.

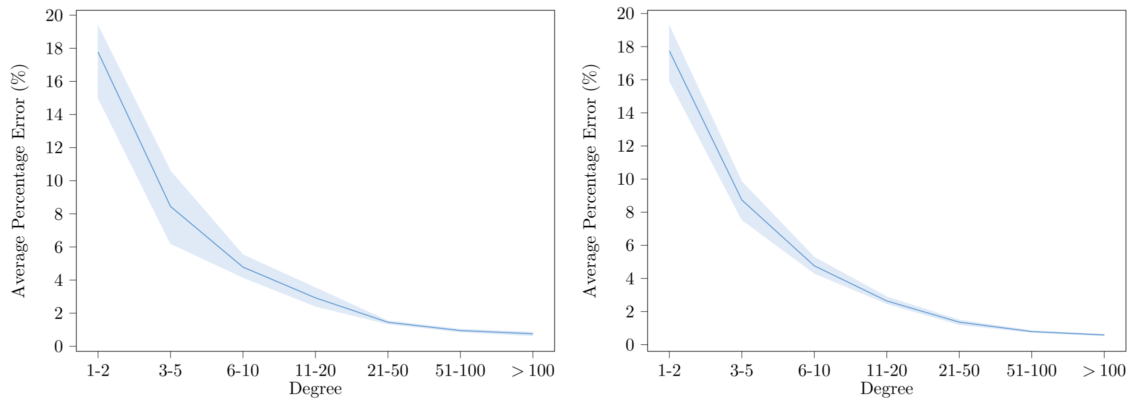
B.5. Experiment Results on Real-world Networks

We randomly generate 10 diffusion instances for *Caltech36* and *Amherst41*, respectively. We list the statistics of the MAPE and WMAPE for both instances in Table 5. It is notable that MAPE and WMAPE of the FPA solution for both networks are stably small.

Table 5 **Statistics of the FPA error for real-world networks**

	MAPE (%)		WMAPE (%)	
	Mean	90% CI	Mean	90% CI
Caltech36	3.29	[2.96,3.61]	1.21	[1.15,1.32]
Amherst41	1.71	[1.63,1.82]	0.75	[0.71,0.80]

Next, in Figure 14, we show the average percentage error of agents with different degree range in 10 repetitions. For both networks, we observe the same trend that the percentage error decreases rapidly when the agent has a large degree. These evidence further justifies our theoretical results from the perspective of the diffusion on a single network.



(a) *Caltech36* (b) *Amherst41*

Figure 14 Average percentage errors for agents with similar degrees in a single network instance (Color shaded area denotes the 90% confidence interval)

Appendix C: Proofs and Experiments in Section 5

C.1. Proofs in Section 5

Proof of Theorem 3: Consider two seed set $S_1 \subseteq S_2 \subseteq V$ and an additional user $w \in V \backslash S_2$, it is sufficient to show that $\boldsymbol{\mu}^*(S_2 + \{w\}) - \boldsymbol{\mu}^*(S_2) \le \boldsymbol{\mu}^*(S_1 + \{w\}) - \boldsymbol{\mu}^*(S_1)$.

We consider the dynamical system version of the constraints (9b) and (9c), that is, $\boldsymbol{\mu}(t) = \boldsymbol{h}(\boldsymbol{\mu}(t-1))$. We can notice that, for different seed sets, the transition function $\boldsymbol{h}$ are not the same. However, for all the users that are not selected as seed users, the transition function for the corresponding element are the same. With a little abuse of notation, in the following proof, we use $\boldsymbol{h}$ to denote the transition function for all users in $V \backslash (S_2 + \{w\})$.

We want to show that at all time steps $t = 1, 2, \cdots$, the inequality $\boldsymbol{\mu}(S_2 + \{w\}, t) - \boldsymbol{\mu}(S_2, t) \le \boldsymbol{\mu}(S_1 + \{w\}, t) - \boldsymbol{\mu}(S_1, t)$ always holds. For user $i \in S_2$, $\mu_i(S_2 + \{w\}, t) - \mu_i(S_2, t) = 0 \le \mu_i(S_1 + \{w\}, t) - \mu_i(S_1, t)$. For user w, $\mu_w(S_2 + \{w\}, t) - \mu_w(S_2, t) = 1 - \mu_w(S_2, t) \le 1 - \mu_w(S_1, t) \le \mu_w(S_1 + \{w\}, t) - \mu_w(S_1, t)$. The above two inequalities hold because of Theorem 2(i). For all the other users in $V \backslash S_2 \cup \{w\}$, we show by induction.

$\boldsymbol{t = 1}$: First of all, $\boldsymbol{\mu}(S_2, 0) \ge \boldsymbol{\mu}(S_1, 0)$ by definition. According to the convexity in Assumption 3, the transition function $\boldsymbol{h}$ is element-wise concave. Therefore, $\boldsymbol{h}(\boldsymbol{\mu}(S_2 + \{w\}, 0)) - \boldsymbol{h}(\boldsymbol{\mu}(S_2, 0)) \le \boldsymbol{h}(\boldsymbol{\mu}(S_1 + \{w\}, 0)) - \boldsymbol{h}(\boldsymbol{\mu}(S_1, 0))$ which implies $\boldsymbol{\mu}(S_2 + \{w\}, 1) - \boldsymbol{\mu}(S_2, 1) \le \boldsymbol{\mu}(S_1 + \{w\}, 1) - \boldsymbol{\mu}(S_1, 1)$.

Assume $\boldsymbol{t = s}$: The induction hypothesis holds such that

$$\boldsymbol{\mu}(S_2 + \{w\}, s) - \boldsymbol{\mu}(S_2, s) \le \boldsymbol{\mu}(S_1 + \{w\}, s) - \boldsymbol{\mu}(S_1, s)$$

Show $\boldsymbol{t = s + 1}$: We have

$$\begin{aligned}
\boldsymbol{\mu}(S_2 + \{w\}, s+1) - \boldsymbol{\mu}(S_2, s+1) &= \boldsymbol{h}(\boldsymbol{\mu}(S_2 + \{w\}, s)) - \boldsymbol{h}(\boldsymbol{\mu}(S_2, s)) \\
&\le \boldsymbol{h}(\boldsymbol{\mu}(S_2, s) + \boldsymbol{\mu}(S_1 + \{w\}, s) - \boldsymbol{\mu}(S_1, s)) - \boldsymbol{h}(\boldsymbol{\mu}(S_2, s)) \\
&\le \boldsymbol{h}(\boldsymbol{\mu}(S_1 + \{w\}, s)) - \boldsymbol{h}(\boldsymbol{\mu}(S_1, s)) \\
&= \boldsymbol{\mu}(S_1 + \{w\}, s+1) - \boldsymbol{\mu}(S_1, s+1)
\end{aligned}$$

where the first inequality comes from Theorem 2(i) and the second inequality comes from Assumption 3.

When t tends to infinity, we get the fixed-point solution $\boldsymbol{\mu}^*(S_2+\{w\})-\boldsymbol{\mu}^*(S_2)\leq \boldsymbol{\mu}^*(S_1+\{w\})-\boldsymbol{\mu}^*(S_1)$, and hence the submodularity is proved. $\square$

Proof of Theorem 4: Let $\pi(\boldsymbol{\mu})=\sum_{i\in V}\left(v_i+\beta\sum_{j\in\mathcal{N}_i}\frac{\mu_j}{n_i}+\ln\frac{1-\mu_i}{\mu_i}\right)\mu_i$. The Hessian matrix of $\pi(\boldsymbol{\mu})$ can be derived as

$$\frac{\partial^2\pi}{\partial\mu_i^2}=-\frac{1}{\alpha}\frac{1}{\mu_i(\mu_i-1)^2}\quad\text{and}\quad\frac{\partial^2\pi}{\partial\mu_i\partial\mu_j}=\frac{1}{\alpha}\mathbb{1}\{j\in\mathcal{N}_i\}\frac{\beta}{n_i}+\frac{1}{\alpha}\mathbb{1}\{i\in\mathcal{N}_j\}\frac{\beta}{n_j}.$$

For the diagonal elements of the Hessian matrix H_π, we can have $-1/[\mu(\mu-1)^2]\leq -6.75$ holds for any $x\in[0,1]$. The inequality is tight when $\mu=1/3$.

For the nondiagonal elements of the Hessian matrix H_π, we can find them related to the structure of network $G(V,E)$. The row normalized adjacency matrix A_{row} is defined as $D^{-1}A$ where D is the diagonal node degree matrix and A is the adjacency matrix. As a result, all rows of A_{row} sum to 1.

Therefore, we can have the Hessian matrix to be

$$H_\pi=\operatorname{diag}\left(\left\{-\frac{1}{\mu_i(\mu_i-1)^2}\right\}_{i\in V}\right)+\beta\left(A_{\text{row}}+A_{\text{row}}^\top\right)\preceq -6.75I+\beta\left(A_{\text{row}}+A_{\text{row}}^\top\right)$$

By Gershgorin circle theorem, we can bound the eigenvalues of A_{row} by $-1\leq\lambda\left(A_{\text{row}}\right)\leq 1$. Since 1 is one of the eigenvalues of A_{row}, we can have $\lambda_{\max}\left(A_{\text{row}}\right)=1$. Therefore, when $\beta\leq 3.375$,

$$\lambda_{\max}\left(-6.75I+\beta\left(A_{\text{row}}+A_{\text{row}}^\top\right)\right)\leq 0$$

which implies that H_π is negative semi-definite. The equality holds when $\beta=3.375$.

As a result, $H_\pi\preceq 0$ if and only if $\beta\leq 3.375$. $\square$

C.2. Comparison between MCMC and the FPA Scheme in Pricing Problem

In order to check the performance of the FPA scheme with regard to the total profit when price is considered, we test over three groups of instances. By fixing the expected number of neighbors to be 10, we generate diffusion instances with random graphs $G(20,0.5)$, $G(100,0.1)$, $G(1000,0.01)$. For each instance, the agent is associated with an intrinsic value i.i.d. sampled from $\mathcal{U}(0,4)$ and an offered price i.i.d. sampled from $\mathcal{U}(0,4)$. We set $\beta=3$ and $\alpha=1$. In Figure 15, we show the distribution of profit difference among all diffusion instance. We notice that the absolute profit difference is small. Furthermore, as the network becomes larger, the performance gap becomes more concentrated. In conclusion, we consider the FPA scheme can achieve almost the same performance as simulation in pricing problem.

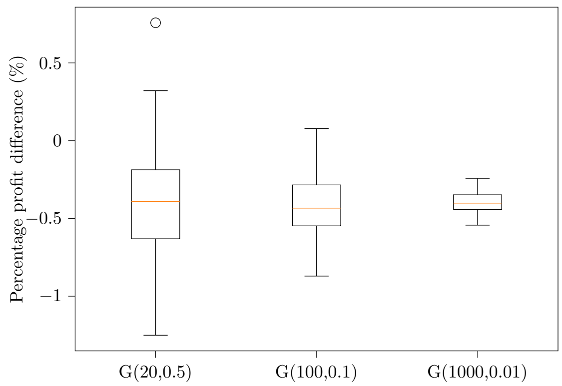


Figure 15 **Profit difference between MCMC and the FPA solution**